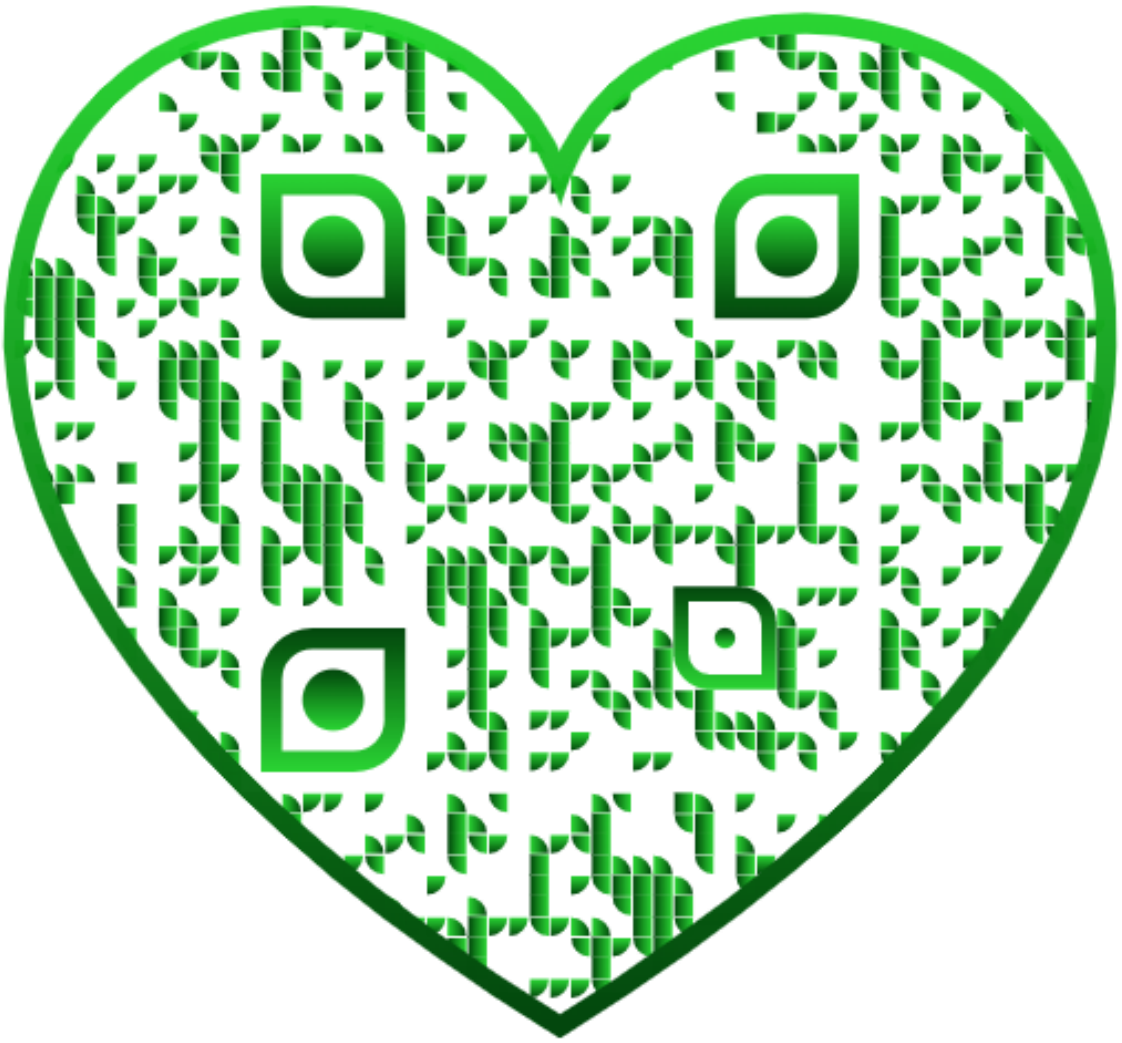
HOW THE UNIVERSE ORGANIZES ITSELF TO CREATE LIFE, CONSCIOUSNESS, AND COSMIC COMPLEXITY

# THE ROMANCE OF REALITY

BOBBY AZARIAN

"What if our cosmic purpose is to create our own cosmic purpose? This book will blow your mind."

—MICHAEL SHERMER, author of *Why Darwin Matters* and *The Moral Arc*



# THE ROMANCE OF REALITY

## HOW THE UNIVERSE ORGANIZES ITSELF TO CREATE LIFE, CONSCIOUSNESS, AND COSMIC COMPLEXITY

BOBBY AZARIAN

BenBella Books, Inc.
Dallas, TX

# CONTENTS

*Introduction*

**PART ONE**
**ORIGINS**

**1** A New Beginning

**2** Energy, Entropy, and the Paradox of Life

**3** Unraveling the Mystery of Life

**4** The Emergence of Life on Earth

**5** Biological Information and Computation

**PART TWO**
**EVOLUTION**

**6** Natural Selection and Knowledge Creation

**7** A Unifying Theory of Reality

**8** The Integrated Evolutionary Synthesis

**9** Hierarchical Emergence

**PART THREE**

**TRANSCENDENCE**

**10** The Mind-Body Mystery

**11** Causal Emergence and Free Will

**12** The Strange Loop That Spawns the Self

**13** Transcendence and Enlightenment

*Acknowledgments*

*Notes*

*Index*

# INTRODUCTION

---

We are a way for the cosmos to know itself.

**—Carl Sagan**

---

It is a thrilling time to be alive—perhaps more so than any other time in human history. Right now, we are beginning to experience a paradigm shift, and it is a profound one. A *paradigm* is a general scientific worldview, and a paradigm shift occurs when new science forces us to adopt a different overall framework and perspective. Such a shift occurred when humanity learned that the earth was not the center of the universe, but instead a seemingly insignificant planet among countless others. Another major shift happened when Charles Darwin's theory of evolution by natural selection explained that all life in the biosphere evolved from a single common ancestor, that is, a single-celled organism.

While the paradigms that these discoveries ushered in taught us a lot about the universe and our origins, they helped shape a worldview that depicted life in the cosmos as accidental and utterly insignificant in the grand scheme of things. This view, which resonated with critics of religion and enemies of superstition, was reinforced by subsequent scientific breakthroughs and the popular philosophical interpretations of these new laws and processes.

For example, around the same time that Darwin was writing *On the Origin of Species*, scientists were formulating the second law of thermodynamics, a principle that created an overwhelmingly bleak cosmic narrative. Not only was the universe's useful energy supply constantly being depleted—like a great engine running out of fuel—a new statistical understanding of the law seemed to imply that the world was continually becoming more disordered and random. If true, it would mean that all forms of complexity and organization, including intelligent life, are doomed to a transient and ultimately insignificant existence, cosmically speaking.

The significance of life took another blow around the mid-twentieth century. The discovery of the DNA molecule confirmed Darwin's big idea: all forms of life in the biosphere, from hamsters to humans, were produced by a blind and mindless mechanical process that could be boiled down to replication with genetic mutation, which inevitably leads to speciation. While Darwin himself was careful not to assume that this was nature's sole evolutionary mechanism, many proponents of his theory did just that. Since in their minds the creation of complex forms required biological evolution, the actual *emergence* of life came to be portrayed as the result of some improbable molecular collision rather than a lawful evolutionary process, like self-organization. In other words, life was a statistical fluke—a "cosmic accident" so unlikely that we should not expect it elsewhere in the universe.

This general sentiment was in line with the dominant scientific ideology of the nineteenth and twentieth centuries, known as *reductionism*, which proposed that reality could be best understood by breaking down all physical phenomena to their simplest parts and processes, so that we may observe the basic behavior of the fundamental constituents of nature in isolation. According to the gospel of reductionism, whenever possible, the social sciences and psychology should be reduced to biology, biology should be reduced to chemistry, and chemistry should be reduced to fundamental physics. Although the reductionist approach has been wildly successful, giving us most of our greatest physical theories, it created the impression that all life-forms, including humans, are nothing more than collections of atoms obligatorily following fixed and arbitrary mechanical trajectories, determined solely by math and not by mind.

As a result, the reductionist approach to science helped popularize the philosophical stance known as *materialism*, which holds that reality only

consists of that which is physical. While this view helped further rid science of supernatural concepts like souls and spirits, classic materialism denied the existence of *apparently* immaterial things, like consciousness, and largely ignored the concepts of energy and information. In doing so, materialism reduced us to zombielike meat machines with no agency, feeling, or inner experience. To most materialists, life and mind are considered to be "epiphenomena" if acknowledged at all, which essentially means they don't matter; they are just *there*. Any feeling of free will we might have when making a decision is merely an illusion. We are not the authors of our actions but passive observers who get constantly fooled by the brain into believing we're not causally impotent.

Collectively, these intellectual advances formed what philosophers call the *reductionist worldview*. This paradigm was not satisfied merely with the removal of god and the soul from the physical picture—it also wanted to purge nature of all traces of purpose or progress. Regarding such important existential questions as "How did we get here?" and "Where are we going?" the reductionist worldview answers "luck" and "probably nowhere." It was supremely rational to adopt this worldview, as that is how things first appear to those who have rightly abandoned explanations for natural phenomena that invoke the supernatural. But the paradigm that is on the horizon leads us to a very different conclusion—one that will radically transform the way we think about the universe and our place in it. In particular, the phenomena that we call life, consciousness, and intelligence will be shown to have deep cosmic significance.

These insights come from considering the roles that energy and information play in the emergence and growth of complexity. Such considerations are the realm of a relatively new academic discipline known as *complexity science*, a unification of the major sciences of our time including, but not limited to, physics, biology, neuroscience, computer science, evolutionary theory, and statistics. These sciences and their methods have been combined by researchers with interdisciplinary backgrounds to form an integrated approach, equally theoretical and experimental, aimed at understanding how nature's dynamical systems emerge and evolve over time. A *dynamical system* is a general term for any system made of a set of interacting components that can explore a variety of structural or functional states, and they can be physical, chemical,

biological, cognitive, social, or technological. Complexity science studies dynamical systems on all scales, from the imperceptibly small to the inconceivably large, including the largest dynamical system of all, the universe itself.

The methods employed by complexity scientists—the stepchildren of the cyberneticists and chaos theorists of the twentieth century—achieve something that the reductionist approach is fundamentally incapable of: They allow us to understand how nature's building blocks spontaneously self-assemble through a synergistic dance that creates wondrous emergent phenomena, like life, mind, and civilization. It turns out that the collective behavior of interacting parts, not simply how they function in isolation, is key to understanding the emergence and evolution of all the fascinating organisms and ecosystems that make up the biosphere. These kinds of dynamical systems are special in their ability to adapt to a changing environment, and they have been suitably named *complex adaptive systems*. We can generally think of life as a form of *adaptive complexity*, to distinguish it from the non-adaptive forms of complexity, like the order we see in structures such as crystals and snowflakes, which is fixed and not functional. Adaptive complexity is a flexible term that allows us to talk about the entire biosphere as an integrated and interconnected network of complex adaptive systems.

Adaptive complexity has often been described as existing at the boundary between order and chaos—the "edge of chaos" it's been called—and it is at this juncture where structure and randomness conspire to create systems that are optimally resilient, flexible, and innovative. Due to the intricate and convoluted nature of complex adaptive systems, from cells to communities, their dynamics could not be adequately understood or predicted until computer modeling reached a certain level of sophistication. Recognizing this fact, the theoretical physicist Stephen Hawking famously declared that the twenty-first century would be "the century of complexity."

From a deeper understanding of how complexity emerges in nature, a new cosmic narrative is being revealed that will change our understanding of our origins and our future. I call it new, but paradigms don't come about overnight. They often spread gradually at first, slowly permeating scientific, philosophical, and intellectual circles until a tipping point is reached, triggering a sudden rise to prominence. Well, that tipping point is on our

doorstep. While this new "big picture" will likely surprise most people, it may seem logical if not obvious to those curious minds who have tried to imagine the future implied by the exponential rate of technological progress. Whether the paradigm I speak of confirms intuitions, surprises you, or triggers a skeptical reaction, it is likely to excite you, because in a strong sense, it is about you.

Through rigorous scientific argument and mechanistic explanation, *The Romance of Reality* will elucidate why it is not just the biosphere but the entire universe that is undergoing an adaptive transformation. The assumption that our world is gradually drifting toward a more disordered, random, and lifeless state is utterly wrong and the result of a fundamental misunderstanding of thermodynamic law. If it is accurate to think of the cosmos as a massive computational machine, it is not one that is winding down. In terms of adaptive complexity, it appears to be just getting started. Through a series of hierarchical emergences—a nested sequence of parts coming together to form ever-greater wholes—the universe is undergoing a grand and majestic self-organizing process, and at this moment in time, in this corner of the universe, we are the stars of the show.

As cosmic evolution proceeds, the world is becoming increasingly organized, increasingly functional, and, because life and consciousness emerge from sufficient complexity and information integration, increasingly sentient. Through the evolution and eventual outward expansion of self-aware beings like ourselves, and their efforts to organize matter into arrangements that support information processing and computation, the universe is, in a very real and literal sense, waking up. It is not waking up independent of us, as in a panpsychic sense, but *through us*, as all the matter that composes life was once inanimate. As cosmologist and science educator Carl Sagan famously put it, "We are a way for the cosmos to know itself." Rather than brushing it off as merely a poetic metaphor, this book takes Sagan's statement seriously and places it within the context of cosmic evolution. In doing so, we see that adaptive complexity has initiated a cosmic awakening process that is only just beginning. Where it ends is to be determined. Exactly how the story goes curiously appears to crucially depend on the actions of intelligent life. In fact, it could depend in some meaningful way *on you*, and *on us*, collectively.

While the laws and constants of physics seem to have carved out this evolutionary trajectory for life in the universe—described by futurists like Google's Ray Kurzweil as a "cosmic destiny," determined not in detail but in rough outline—the hard truth is that existential success is in no way guaranteed for *Homo sapiens*. There is no particular law or force of nature preventing our civilization from failing. Progress occurs not because it is being driven by some mystical force, but because life learns from continual failure. Natural selection is nature's algorithm for error correction. If we are not wise, we could be the errors that get corrected. In that case, the next civilization or species will have their go, and if they do not repeat our mistakes, they will advance beyond where we were at the height of our glory. This book aims to show that knowledge in the biosphere will accumulate under either scenario.

But we don't have to fail. The fact that humans truly are, as this book will argue, autonomous agents with *free will*—formalized as *causal power*—means that it is up to us to decide between extinction and transcendence. We have every incentive to work toward the latter and away from the former. Through the collective efforts of humanity, by way of our intellectual, cultural, and technological progress, we can continue to assist the cosmos in its great awakening process. Promoting an awareness of our emergent cosmic purpose could facilitate exponential social, economic, and technological progress, allowing us to transcend our intrinsic limitations and the biosphere to extend itself outward into the heavens.

Of course, the new cosmic narrative should not be taken on faith. Any claims so magnificent should, as a rule, be greeted with thorough skepticism. A paradigm shift of this magnitude requires a satisfying physical account of the evidence supporting such a radical departure from general scientific consensus. But that account does indeed exist, and the pages to come will describe the exciting new theories, experimental findings, and physical processes that substantiate these remarkable assertions.

The major emergences in the self-organization of the universe will be explained mechanistically, so that we may see precisely how and why adaptive complexity and the knowledge it embodies grows inevitably and without bound, as a consequence of the laws of physics and the evolutionary dynamics that emerge from the constraints they impose on

matter in motion. In our quest to understand cosmic evolution, we will arrive at a "theory of everything" that we may call a *unifying theory of reality*.

This ambitious theory attempts to solve the greatest remaining mysteries of science. The infamous "hard problem of consciousness," the puzzle of free will, and the mystery of increasing cosmic complexity in an increasingly entropic universe all begin to unravel as the unifying theory dissolves the paradoxes created by the unjustified assumptions of the reductionist worldview and exposes the language traps that have prevented us from making intellectual progress for so long. By introducing a new evolutionary synthesis—*the integrated evolutionary synthesis*—the unifying theory of reality bridges the gap between the quantum and cosmological with principles from evolutionary biology.

This book is an invitation to a cosmic journey dedicated to understanding the universe, how it is waking up, and what that means for us as human beings. Part One of this book, "Origins," is about the emergence of life on Earth. To understand this event, we must become familiar with the basic concepts of complexity science and cybernetics, such as self-organization, phase transitions, attractors, and feedback loops. Part Two, "Evolution," is about the emergence and evolution of intelligent life in the cosmos. Evolutionary epistemology, universal Darwinism, and an emerging paradigm called *universal Bayesianism* will allow us to integrate the concepts from Part One into a unified theory of systems. Part Three, "Transcendence," covers consciousness, free will, and the ultimate fate of life in the universe.

By the end of the trip, we will have arrived at a new scientific and spiritual worldview—called *poetic meta-naturalism*—that challenges everything we thought we knew about the world. According to this philosophy, reality is inherently creative because it is always producing novel patterns that are "rhymes" on the patterns of the past. The knowledge this new worldview offers about the power of knowledge may be the only thing that can save our specific civilization from self-destruction. The existential game clock is ticking, so let's get started.

# PART ONE

# ORIGINS

*We curiously find ourselves in a pro-life universe*
*But that universe is not conscious, at least not at first*

# 1

# A New Beginning

> The origin of life must be a highly probable affair. As soon as conditions permit, up it pops!
>
> **—Carl Sagan**

The great nineteenth-century English biologist Thomas Huxley, known as "Darwin's bulldog" for his advocacy of evolutionary theory, believed that "the question of questions for mankind—the problem which underlies all others, and is more deeply interesting than any other—is the ascertainment of the place which Man occupies in nature and of his relations to the universe of things."[1]

In other words, how do we fit into the grand cosmic picture? The aim of this book is to answer this fundamental question. The overarching thesis is that we live in a computational universe that is continuously evolving into an increasingly complex, functional, and sentient state. This means that humans are neither a cosmic accident nor the end goal of evolution. Instead, we are an intermediary step on the cosmic evolutionary ladder of becoming.

This emerging view is supported by recent advances in origin-of-life research, which suggest that the appearance of biological systems on Earth roughly four billion years ago was not an improbable occurrence, as

previously thought, but an *inevitable* one. The new consensus marks the start of a radical shift away from the reductionist worldview, which sees life as accidental, fragile, rare, and transient, toward an evolutionary model of the world where the open-ended growth of order, complexity, and knowledge occurs inevitably as a result of natural mechanistic processes.

Upon first consideration, arguing over whether life in the universe was an accident or an inevitability might sound like a pointless exercise in semantics. One is tempted to say either life happened or it didn't—end of story. But it is more nuanced than that. Remember that life is a complex adaptive system. Whether the emergence of adaptive complexity was improbable or inevitable is a scientific question that has a precise and discoverable answer. To call something accidental implies that the event was determined only by chance. This means that a slightly different chain of events would be expected to produce a different outcome. These chance-based outcomes are called "contingencies," and they are impossible to predict in advance. On the other hand, we can say life was inevitable if the laws of physics create dynamics that constrain the random and chaotic motion of particles in ways that guarantee biological systems emerge where conditions permit.

The question is as practical as it is philosophical, since the two distinct views make incompatible scientific predictions. For example, they arrive at wildly different conclusions about the prevalence of extraterrestrial life, a topic of great public interest. If life is highly improbable, we are likely to be alone in the universe. If the emergence of life is inevitable given a certain set of common physical conditions, then we are likely to have many cosmic companions spread throughout space on planets in "habitable zones." Whether they'd likely be intelligent is a separate matter, but an equally important one to the new cosmic narrative, so we will come back to it in Part Two.

The process by which life arose from non-life is known as *abiogenesis*. In the coming chapters, we will be introduced to a new model of abiogenesis that suggests this extraordinary emergence was the inevitable beginning of a preordained cosmic awakening process, determined not in perfect detail but in broad outline. The origin of life is the story of how adaptive complexity first gets going. But to fully appreciate how this

happens mechanistically and why this version of abiogenesis is so different from the old one, some historical context is needed.

As we have learned, the reductionist worldview generally portrays life on Earth as a cosmic accident. If we could rewind the tape of time and play it again with even the slightest change, it's unlikely that life would have emerged on our planet, or for that matter, in the entire universe.

The Nobel Prize–winning French biologist Jacques Monod—a staunch skeptic and righteous enemy of organized religion—poetically summed up this worldview in his influential book *Chance and Necessity*, published in 1970, when he said, "The ancient covenant is in pieces; man knows at last that he is alone in the universe's unfeeling immensity, out of which he emerged only by chance. His destiny is nowhere spelled out, nor is his duty." It was Monod's passionate and uncompromising belief that "the universe was not pregnant with life."

Given their knowledge of the vastness of our universe, what made twentieth-century scientists like Monod so sure that life was a statistical anomaly? The truth is, it was nothing more than an assumption. At the time, it was widely believed that biological evolution was nature's only mechanism for generating complexity, and any system that came about prior to that process, they reasoned, simply had to be a product of chance events. Thus, the origin of life was described as the result of "chance assembly," which is just a fancy term for cosmic luck.

## Darwin Explained How Life Evolves, Not How It Emerged

Despite its ability to explain so much about the amazing diversity of the biosphere, there are some things that Darwin's original theory simply can't explain about the adaptive complexity we see around us, such as its origin.

Biological evolution through natural selection is an elegant concept that can be summed up quite easily, thanks to our modern understanding of genetics. Living organisms are self-replicating systems that use a set of embedded instructions (encoded in DNA) to make copies of themselves. Cells divide and organisms reproduce. This is how life persists on Earth

over incredibly long time scales, even though individual organisms are quickly coming and going. Life, among other things, is copy-making machinery.

However, the copying process never occurs perfectly, and errors are unavoidable. These errors give rise to genetic mutations that generate a wonderfully diverse set of new and modified organisms, different from the parents and different from one another, with similar but varying physical traits, forms, functions, and behaviors.

In order to replicate and pass on new characteristics to the next generation, an organism must first be able to survive long enough to do so in a competitive and often harsh environment. The traits and functions that better allow organisms to survive and reproduce are the ones that will persist in the population, while the others, and the organisms they belong to, die out. In this way, environmental pressures prune the diversity created by genetic mutation, much like a bonsai tree is pruned by a gardener. This is what is meant by "natural selection." Nature selects the most "fit" survival features from the possibilities offered up by the genetic lottery game. As a result of being sculpted by the forces of nature, over time organisms become better adapted to the environment they inhabit. Simple niches tend to produce relatively simple organisms, while niches that present a more complex set of challenges tend to produce more complex organisms.

To many, it seemed that the mystery of complexity was entirely solved by the theory of biological evolution. Essentially, all living order, from daffodils to dolphins, can be reduced to one general mechanism—natural selection acting on genetic mutation. The problem with this story is that it begs the question: How did the first living organism, a self-replicating system, come about in the first place? Even RNA, the DNA molecule's simpler cousin, is an extremely complicated structure rich with information and chemical complexity. Not only that, self-replication is a mechanical process that must be powered by energy. So, in addition to genetic material, a biological system requires an elaborate chemical process called *metabolism*, which relies on a structured chemical network carrying out a complex set of reactions. Biological evolution can only explain how these features evolve, not how they emerged in the beginning.

With no conceivable mechanism to explain how a collection of inanimate particles could have assembled themselves into a functioning

cell, scientists of the time were forced to attribute abiogenesis to one giant stroke of statistical luck, rather than some yet-to-be-discovered proto-Darwinian evolutionary process. This view, based on the idea of chance assembly, became known as *the chance hypothesis*.

## Life Is Not a Fluke

According to the chance hypothesis, against all odds, through some improbable collision in a "primordial soup" on the ancient Earth, the right molecules fell together in just the way needed to create a living organism: a great "frozen accident," it has been called. While those who formulated this hypothesis—the codiscoverer of DNA, Francis Crick, among them—had only a vague idea of how complicated the simplest system capable of self-replicating and metabolizing would have to be, even they knew that such an event would be incredibly rare. But with no other mechanistic theory to account for abiogenesis, they settled on chance.

While Crick playfully called life a "happy accident" earlier in his career, he was keenly aware of the problems with such a theory. On this point he once remarked, "An honest man, armed with all the knowledge available to us now, could only state that in some sense, the origin of life appears at the moment to be almost a miracle, so many are the conditions which would have had to have been satisfied to get it going."[2]

For this hypothetical molecular collision to have taken place given the kinds of molecules in the presumed primordial soup, nature would have had to defy the odds. A number of scientists, puzzled by the problem, attempted to calculate the probability of such an event and found it to be unlikely to have occurred in the projected lifespan of the cosmos. Fred Hoyle, a famous British astrophysicist who had a hard time swallowing the story, expressed his skepticism in his book *The Intelligent Universe* with a devastating analogy: "The chance that higher life forms might have emerged in this way is comparable to the chance that a tornado sweeping through a junkyard might assemble a Boeing 747 from the materials therein."

But that's not the only reason to question the "cosmic accident" origin story. Life appears to have emerged on Earth just about as soon as

conditions allowed, which would be a massive coincidence if its emergence was in fact improbable and a product of pure chance. If the latter were the case, one would expect life to have appeared after some arbitrary amount of time once the planet became habitable—literally any other time besides almost immediately. Christian creationists were quick to capitalize on this fact. In their minds, since chance couldn't adequately explain the emergence of life, it must have been a supernatural act. While this theory is unsatisfactory because it attributes life to a literal miracle, the chance hypothesis is unsatisfactory because it relies on a statistical miracle.

Fortunately, a third option emerged in the twentieth century, and it is one that a growing number of scientists find more appealing than the chance hypothesis or divine intervention. What if the emergence of life was the inevitable result of a universe that naturally and ceaselessly generates complexity and information? If that were the case, it would be more accurate to describe life as a regularity rather than an anomaly—not a fluke but an imperative.

Sagan was an early proponent of this view, which gave astronomers and alien enthusiasts new hope for finding extraterrestrial life. "The origin of life must be a highly probable affair," he exclaimed, "as soon as conditions permit, up it pops!"[3] This position was further articulated by Christian de Duve, the Belgian biochemist who won the Nobel Prize in 1974. He believed that Monod got it wrong: "If you equate the probability of the birth of a single bacteria cell to chance assembly of its atoms, eternity will not suffice to produce one . . . Faced with the enormous sum of 'lucky draws' behind the success of the evolutionary game, one may legitimately wonder to what extent this success is actually written into the fabric of the universe."[4]

In his cogent book *Vital Dust*, de Duve responded to Monod's claim that the universe wasn't pregnant with life, nor the biosphere with man, with the simple rebuttal, "You are wrong. They were." Of course, de Duve did not mean that *Homo sapiens in particular* were destined to arise. Contingency certainly has its role in biology, which he acknowledged in all his writings. But he did believe that the laws of physics and the evolutionary dynamics that emerge from them generate increasingly sophisticated forms of

biological organization and complexity. This in turn inevitably produces life and mind in the cosmos.

If such a view is correct, then it would be a philosophical game changer, with implications that could best be described as spiritual. Stuart Kauffman, a complexity scientist at the Santa Fe Institute who has arguably done the most to bring the "inevitable life" theory into the mainstream of science, puts it romantically in his now-classic 1996 book, *At Home in the Universe*: "If we are, in ways we do not yet see, natural expressions of matter and energy coupled together in non-equilibrium systems, if life in its abundance were bound to arise, not as an incalculably improbable accident, but as an expected fulfillment of the natural order, then we truly are at home in the universe."

But the claim that life was written into the laws of nature from the beginning is an extraordinary one, and as Sagan put it, "Extraordinary claims require extraordinary evidence." If life was indeed written into the laws of nature from the very beginning, why is Earth the only planet where we've observed this supposed regularity? How can we say anything about the likelihood of life elsewhere if we are restricted to a measly sample size of one?

Furthermore, if life didn't form through a singular chance event, what is the physical mechanism that explains how a group of inanimate particles could systematically transition into a biological system with function, and under what natural conditions should we expect it to happen? Is it really possible to show that life was, in a physical and statistical sense, *destined* to emerge in the universe, on planets like Earth? The growing body of evidence supporting a universe that innately churns out biology is becoming overwhelming.

## Energy as the Fundamental Organizer

When we ask how life emerged, what we are really asking is how disorganized matter suddenly became organized. In our everyday experience, spontaneous organization doesn't just happen. Things don't organize themselves, and unless someone intervenes, rooms get messier, buildings erode, and intricate structures gradually disintegrate. So, how is it

possible to have organization before there were any beings around to build or maintain it?

The simple answer to that complicated question is *energy*. Energy is nature's fundamental organizing force, and a universe rich with flowing energy is a cosmos primed for complexification. As alluded to in the introduction, it is the concentrated flow of energy through matter that arranges nonliving elements into a configuration that is "alive." By promoting self-organization, energy flux gradually converts a clump of mindless molecules into information-processing machinery with *agency*—meaning the self-drive and purposeful action that we associate with life.

Agency is the characteristic that is absent from the world of nonliving things, like rocks and chairs, which don't move unless they are pushed. Agency used to be attributed to a mystical force driving life toward some greater goal, but as we will learn by the end of Part One, it is enabled by the information processing performed by biological machinery. Because they possess agency, living things are also known as *agents*. The flow of energy builds agents from inanimate matter by organizing chemical systems into autonomous computational machines that gather information about the world and use that knowledge to perpetuate and spread life further.

While this process might sound mystical, we are talking about energy as defined by conventional science. It is detectable and measurable. The order-promoting quality of flowing energy is entirely natural, but surprisingly, it is not that conceptually different from the old notion of an élan vital, a "vital force" that animates the inanimate. You could say that energy organizes matter into life, and life organizes information into knowledge. And it is knowledge that makes the universe comprehensible to intelligent life. Or, from a less anthropocentric perspective, it is knowledge that allows the universe to comprehend itself, *through* intelligent life. We must remember that we are not distinct from nature but a functional manifestation of it.

To understand how this life-generating, knowledge-creating process works at an intellectually satisfying level, we must first gain a little familiarity with the branch of physics that deals with energy flow, known as *thermodynamics*. This might be the most challenging part of the cosmic story, but it will be well worth the effort, as it will explain the mystery of how life came into existence, and it will arm us with the conceptual tools

needed to understand the central role it is poised to play in the process of cosmic evolution.

Specifically, we must confront the almighty second law of thermodynamics, which on the surface would appear to contradict the idea that complexity—biological or otherwise—can persist and grow without limit. If that were so, it would make life a transient phenomenon with no cosmic significance—"chemical scum," as Stephen Hawking once put it. Interestingly, it turns out that this fundamental law presents no conflict with the idea of a universe that becomes more complex over time. In fact, the second law of thermodynamics is essential to understanding why complex adaptive systems like life naturally emerge and proliferate.

# 2

# Energy, Entropy, and the Paradox of Life

---

> Scientists have often been baffled by the existence of spontaneous order in the universe. The laws of thermodynamics seem to dictate the opposite, that nature should inexorably degenerate toward a state of greater disorder, greater entropy. Yet all around us we see magnificent structures—galaxies, cells, ecosystems, human beings—that have all somehow managed to assemble themselves.
>
> **—Steven Strogatz, *Sync: The Emerging Science of Spontaneous Order***

---

Some billions of years after the Big Bang birthed it, the universe suddenly awoke from its long slumber. That magic moment occurred when life spontaneously sprang forth from a lifeless planet. To understand how and why this event happened on Earth, and how common life might be in the universe, we must dip our toes into the branch of science known as thermodynamics, which is not just a crucial part of both physics and

chemistry, but as we will learn, biology as well. Of particular interest is the principle of nature known as the second law of thermodynamics.

Before we get into the nuts and bolts of the messy second law, let's start with its simple sibling, the first law of thermodynamics. The first law states that the total amount of energy in an isolated system is constant and unchanging. It can be transformed from one form to another, but it cannot be created or destroyed. An isolated system is one that has no matter or energy coming in or going out. Truly isolated systems are rare if not nonexistent in nature because energy is flowing all around us, radiating outward from stars and planets with cores that generate heat. A basic example of an isolated system is an insulated thermos that keeps coffee hot by trapping the heat inside. Whether the universe itself is an isolated system is a complicated question, so we will come back to it later.

## Entropy: A Complex and Confusing Term

Although the first law is easy enough to understand, the second law of thermodynamics is more cryptic. It states that the total amount of entropy in an isolated system must always increase with time. Many physicists consider the second law to be the most ironclad law in the universe, so this statement deserves some thought and reflection. But what is meant exactly by the nebulous and strangely intimidating term *entropy*?

Unfortunately, the answer is not so cut and dried. Most popular science books, such as Stephen Hawking's classic *A Brief History of Time*, teach us that entropy means "disorder." But this definition can be dangerously misleading when one attempts to apply the second law to systems that aren't isolated, or to the universe as a whole. In a cosmic context, equating entropy with disorder—a word commonly understood to mean roughly the opposite of organization, structure, and complexity—can lead to terribly inaccurate cosmic conclusions.

Unfortunately, the mistake is common among science communicators, who often claim the second law demands that the universe grows more disordered and less patterned with time. Since life *is* biological order, this interpretation of the second law predicts dire consequences for humanity. If the world must always grow increasingly disordered, then on average it

should become more lifeless and less structured over time, until all traces of organization dissolve into a formless and meaningless sea of random microscopic fluctuations. Such a scenario reduces life in the universe to a temporary, trivial, and inconsequential phenomenon—not a very cheery picture of the future. Luckily, it is not the correct one!

The crucial problem with this cosmic narrative is that an increase in the total amount of entropy in the universe does not correspond to an overall decrease in the amount of organization that exists. In fact, biological, cultural, and technological order can grow alongside entropy, and not just in the short term. To understand why this is true, and to untangle all the confusion surrounding the second law, we must distinguish between two different types of entropy (there are many more but we will start with these). For that, a brief history lesson is required.

The concept underlying the second law was developed in the nineteenth century by a handful of especially clever European scientists, most notably Sadi Carnot and Rudolf Clausius, initially for practical reasons rather than theoretical ones. They were searching for the most efficient way to convert energy from heat flow into mechanical energy that could power machines. The steam engine, which had been around for some time, was proof that energy could be extracted from heat flow to do "work," meaning something useful, like moving, lifting, or accelerating a body of mass by exerting a force on it. Work can't be done unless it is driven by a physical force, which always requires energy. The work done by steam engines is moving trains along tracks. This was big business in the 1800s, as trains opened up new possibilities for trade and allowed mass migration from rural areas into industrialized cities.

What Carnot astutely noticed was that if there is a temperature difference between two bodies in contact, heat will spontaneously flow from the hotter to the colder body, until both bodies reach a mutual uniform temperature, a state known as *thermodynamic equilibrium*. Most of us are familiar with this phenomenon, which we experience whenever our warm cup of coffee, or our cozy hot bath, interacts with the colder surrounding air and cools to room temperature. It may be helpful to think of heat flowing from a hot system to a cold one as moving "downhill," which is the natural direction of movement, according to Newtonian physics.

Why does this happen? You might be familiar with the old adage that says, “Nature abhors a vacuum.” Well, it has also been said that “nature abhors a gradient.” A gradient is the difference between two interacting systems that creates instability, whether it be a difference in temperature, pressure, chemical concentration, or electrical charge. If such a difference exists, there will be spontaneous flow from one system to the other until that difference, or the *gradient*, is eliminated, and a stable and inert state of equilibrium is achieved. This happens automatically because nature is simply intolerant of gradients.

So why was this of interest to nineteenth-century engineers? The answer is pretty spectacular. When there is a temperature gradient, the heat that flows to eliminate the gradient creates a physical force that can be harnessed to do work. The greater the temperature difference between two bodies, the steeper the slope of the gradient, and the greater the force of heat flow will be to reduce that difference and restore thermodynamic stability. This is why hot beverages cool off faster when you put them in a refrigerator. The temperature difference between the hot item and its refrigerated surroundings is much greater than in a room-temperature setting.

Engineers who understood this principle realized that a rapidly heated substance would expand with a force powerful enough to do some serious work, if properly harnessed. With a large enough temperature difference between the heated substance and its cooled-down surroundings, the force of heat flow could power an engine. The pressure created by expanding steam could perform work through cycles of rapid heating and cooling.

While an understanding of how to exploit this tendency of nature provided humanity with a powerful new technology, Carnot noticed the conversion of heat energy into mechanical energy was a process that could never be 100 percent efficient. No matter how carefully the transfer was carried out, some of the useful energy would always be lost to the surroundings as waste heat through what physicists call *dissipation*. This term is almost as intimidating as entropy, and it’s equally important to understanding life’s emergence, so we should get comfortable with it right now.

For energy to be dissipated simply means that it gets uniformly dispersed into the surroundings, scattered in such a way that it can never

again be harnessed to do the same amount of work. In other words, dissipated energy is energy that is no longer useful, because the effort required to gather it up would cost more energy than you'd get out of it. A familiar example of energy dissipation is the body heat that we as humans —complex biological systems—constantly give off. Another example is the heat generated by your laptop computer as it computes. A swinging pendulum in a grandfather clock continually dissipates a small amount of energy due to constant friction with the air, which is why it eventually must come to a stop. Every mechanical process that happens in the universe dissipates some amount of useful energy by producing heat. This fact is the basis for the second law of thermodynamics, and it is the reason why the laws of physics do not allow perpetual-motion machines.

This inevitable dissipation has a serious implication for the universe. It means that the universe's supply of useful energy—the "cosmic fuel" that is required for work of any kind—is constantly being depleted as temperature differences are eliminated, energy is dissipated, and heat is generated. While the total amount of energy in the universe presumably stays the same, in accord with the first law of thermodynamics, the *useful* energy is gradually being converted into *useless* energy, in accord with the second law. Useful energy is called "free energy" and useless energy is called "entropy." Stated more precisely, entropy is a mathematical term representing *the quantity of energy no longer available for work*. Since we've already established that there is more than one type of entropy in physics, let's call this kind of heat-related entropy *thermal entropy*. If we think of free energy as cosmic fuel, then thermal entropy is cosmic exhaust.

As you can see, there has been no mention of order or disorder so far. With this definition of entropy, the second law can be stated to apply to the universe, assuming it is an isolated system, in two different but equivalent ways. Due to unavoidable energy dissipation as a result of mechanical processes:

> *The total amount of free energy in the universe must decrease with time.*
> *The total amount of thermal entropy in the universe must increase with time.*

Of the cosmos, this is all that the second law of Carnot and Clausius demands. Contrary to popular belief, it does not in any way forbid life from spreading through the universe and saturating it with structural and functional organization. The only thing that must become gradually "disordered" over time is nature's energy supply, not its large-scale structure. Until the entire free-energy supply of the universe is completely exhausted through dissipation—culminating in an inert state of cosmic thermodynamic equilibrium known as *heat death*—intelligent life with sufficiently advanced technology can in theory persist and expand without limit. There are even some compelling reasons to believe that the universe will *never* reach thermodynamic equilibrium. We'll explore this concept in more detail in the final chapter of Part Three, when we discuss the ultimate fate of the universe.

So where did this notion of entropy as disorder come from? Why did much of the scientific community come to believe that the open-ended growth of life and complexity was prohibited by the second law of thermodynamics? A genius named Boltzmann is largely to blame.

## Boltzmann Entropy: A Measure of Statistical Disorder

As strange as it might sound to us now, just a couple hundred years ago, the idea that physical things are composed of tiny particles called atoms was not universally accepted, even in the physics community. But in the second half of the nineteenth century, support for atomic theory grew rapidly, and physicists began looking for microscale explanations for all previously explained phenomena—since they were considered to be more fundamental by the reigning reductionist paradigm. Soon, thermodynamic laws would be redefined in terms of what was happening in systems at the atomic or molecular level, where tiny particles careen about chaotically and ceaselessly. At bottom, nature is noisy, as its fundamental constituents are constantly fluctuating and colliding.

Building on the prior work of Clausius and James Clerk Maxwell (the Scottish mathematician famous for his theory of electromagnetism), an

Austrian physicist named Ludwig Boltzmann set out to explain the second law, namely the natural tendency for heat to disperse and energy to dissipate, as a result of the statistical behavior of vast numbers of molecules assumed to be moving according to simple laws of mechanics. Boltzmann was inspired to think microscopically by the recent discovery that the kinetic energy of a gas was a direct consequence of how fast its individual molecules were moving. Could the second law be explained by thinking about energy flows in terms of the motions of molecules and their interactions?

He reasoned that if heat energy was nothing but increased molecular motion, then its dissipation must involve a gradual diffusion and dampening of this excited motion over time, which he suspected could be related to random collisions between neighboring molecules. To explore this model of energy dispersion—or entropy creation, since they are two sides of the same coin—he wisely chose a hypothetical system that was infinitely less complex than the universe: an ideal gas in a closed container. To run Boltzmann's thought experiment in your own mind, imagine being able to shrink yourself down to a microscopic size until you can see the behavior of each individual gas molecule. What do you see?

Boltzmann envisioned a bunch of molecules quickly zipping about in all directions, frequently bumping into one another like billiard balls on a pool table, transferring momentum around and spreading each other out in space. Because each molecule's kinetic energy is proportional to its speed, as the molecules randomly collide, the faster ones naturally slow down while the slower ones speed up, until eventually the speed of all become roughly the same. Without any energy differences between molecules, there are no temperature gradients, and no heat flow that can be harnessed to do work. According to this explanation, an isolated system of gas molecules will inevitably approach thermodynamic equilibrium because of the effects of countless unseen molecular interactions.

## Microstates Versus Macrostates

While the exact *microstate* of any system—meaning the exact position and velocity of each individual molecule at a particular time—could not be

measured or calculated precisely for obvious practical reasons, Boltzmann realized that the average collective behavior of the entire ensemble of molecules could be predicted with statistical mathematics.

Unlike microstates, the collective properties of a many-particle system can be easily observed and measured. When measuring the temperature or pressure of a gas, for example, we are really measuring an average value of a collection of molecules, or a *macrostate*. For any particular system of molecules, there are typically many different microstates that correspond to a single macrostate. In other words, there are many different ways to arrange the individual molecules that produce energetically equivalent collective patterns of arrangement. Famously, Boltzmann showed that states of higher entropy have more microstates that all correspond to a single macrostate. Here is where it becomes useful to picture entropy as disorder. Think of it this way: there are many more ways for your room to be a complete mess (disordered) than there are ways of being neatly organized.

Boltzmann described a system evolving toward thermodynamic equilibrium as becoming more "disordered," because no matter how the particles in the system were arranged initially—perhaps the faster moving molecules were neatly bunched up in one corner of the container—the collective configuration would inevitably drift toward a uniform spatial distribution that was devoid of any patterns or discernible structure. For this reason, Boltzmann considered the state of thermodynamic equilibrium and maximum entropy a state of "maximum disorder." You can also think of it as a state where molecular mixing has reached a maximum.

This order-to-disorder transition might at first seem unintelligibly abstract, so let's illustrate the logic with some real-world examples. Let's begin by returning to the billiard ball metaphor. At the beginning of a pool game, the balls are organized neatly in a triangle formation, until the game starts and a player breaks that formation up by striking it with the cue ball. As the configuration evolves over time, the balls become increasingly spread out and disorganized. If there were no pockets for them to roll into, the arrangement of the balls would move closer and closer to an evenly spread-out spatial distribution. There will always be slight deviations away from a state of perfectly equal distribution, but this general configuration will be essentially preserved once it is reached. This more-or-less completely spread-out state represents equilibrium.

Now let's consider some real-world examples that involve the dispersion of heat. When you add some hot water to a cool bath, or some hot coffee to a lukewarm cup, the added liquid is first concentrated in a local blob, and the extractable energy associated with the hot area is considered to be "ordered." As the excited molecules mix and interact with their less excited neighbors, the localized kinetic energy systematically spreads out, temperature gradients dissipate, and the initial energetic order gradually dissolves into entropy. By the time equilibrium is reached, mixing has erased all evidence that the heat blob ever existed. For this reason, the increase in disorder is also associated with a *loss of information* from the perspective of the observer. As mixing proceeds, one has less certainty, or greater ignorance, about the exact microstate the system is in, because *high-entropy macrostates have more equivalent microstates*, or possible arrangements. This connection between entropy and ignorance will become important when we discuss information in Chapter Five, and introduce a third type of entropy: *information entropy*.

## The Birth of Statistical Mechanics

By attempting to explain the second law of thermodynamics using the branch of mathematics known as probability theory, Maxwell, Boltzmann, and the American physicist Josiah Gibbs invented the influential field of physics known as *statistical mechanics*. As the field grew in influence, the term "entropy" became less associated with its original definition—a measure of dissipated, useless energy—and synonymous with the structural or configurational disorder described by Boltzmann. It wasn't long before the second law began to be associated with examples that did not involve any heat flow or energy dissipation, and thus no increase in thermal entropy. In these popular illustrations, the increase in disorder has nothing to do with actual thermodynamics, only statistics.

Since Boltzmann's entropy measure is not about the dispersion of heat per se, but the evolving spatial configuration of a many-particle system's components, we may call it *statistical entropy*, or *configurational entropy*. Although an evolution toward higher configurational entropy can be used to

explain the dissipation of energy gradients, the measure is a mathematical abstraction that has much broader application.

For example, imagine you have a bag full of Skittles that are grouped by color (a highly organized state). The Skittles are analogous to the gas molecules in the container. Now imagine shaking up the bag to simulate random molecular motion. Naturally, the orderly arrangement of the Skittles will become progressively disorganized, until the candies are completely mixed up and there are no longer any discernible patterns of color—a perfectly messy state that represents equilibrium. This inevitable mixing happens for a simple statistical reason. There are many, many more ways for the Skittles to be arranged in a mess than there are orderly looking configurations. Therefore, from sheer probability alone, we should expect any randomly evolving system to go from ordered to disordered over time. This is what happens when you toss a salad and why no amount of shaking is going to untoss it. While in this example statistical entropy increases significantly as the colored candies mix, there is probably only a small increase in thermal entropy, which corresponds to the amount of energy that is dissipated as the candies collide.

A more extreme example of diverging definitions of entropy can be demonstrated with a deck of playing cards. It is easy to see why there are only a small number of card configurations that appear orderly compared to disorderly arrangements. Shuffling a perfectly ordered deck won't do much to increase the deck's thermal entropy, but it will drastically increase the statistical entropy of the deck's configuration.

Although we can see how they are related, thermal entropy and configurational entropy are two distinct concepts that yield subtly different interpretations of the second law of thermodynamics. This made it incredibly easy for nonexperts and physicists alike to conflate the two ideas. It wasn't long before eager scientists and philosophers tried to apply Boltzmann's version of the second law to the universe as a whole, ushering in a new age of cosmic pessimism. Now, not only was the free energy supply of the cosmos being gradually depleted, which was concerning enough, the existing cosmic organization was on an immutable evolutionary trajectory toward increasing decay and disintegration. For humans, the statistical tendency toward increasingly disordered arrangements meant imminent death and doom.

# The Way to Evade a Disordered Fate

Fortunately, this gloomy scenario doesn't match up with reality. Around 3.8 billion years ago, biological organization emerged on at least one planet, and rather than growing more disordered and featureless, the surface of Earth has become increasingly ordered, complex, and functional. So has the cosmos as a whole, by almost any measure. In addition to stellar, planetary, biological, and ecological order, intelligent life has created cultural and technological order in the form of buildings, machines, cities, economies, and information networks—all of which seem to be growing in number and in complexity with no signs of slowing down. These expanding structures are obviously not just the result of transient statistical fluctuations.

So, how do we explain this apparent paradox? How can organization persist and grow if an increasingly disordered state of the system is so much more likely? The lesson to be learned here is that Boltzmann's statistical version of the second law, which states that configurational disorder must increase, only applies to systems that are isolated. For such systems, the inert state of pure molecular chaos is inevitable, as is structural breakdown, because the maintenance of organization represents physical work, and that continual work costs energy. But systems that aren't isolated from external energy flows, called *open systems*, do not have to play by Boltzmann's statistical rules.

These systems are open to energy flows coming in from the environment, which can keep them in a low-entropy state that is far away from thermodynamic equilibrium. Physicists call these open systems *non-equilibrium systems*, and if they are sufficiently far from equilibrium, they are described as *far-from-equilibrium systems*. As long as energy flows are coming in or can somehow be extracted from the environment, a system can resist the otherwise natural drift toward randomness and structural decay.

An example of an open system that is far from equilibrium is the planet Earth, because it constantly receives energy from the light produced by the sun. The most interesting open systems, those furthest from the dreadfully boring equilibrium state, are the biological systems that we call organisms.

For a long time after Boltzmann, physicists thought living organisms were just too complicated to be described thermodynamically, and this was

a major reason why life's emergence has remained a total mystery for so long. But difficult challenges inspire scientists to come up with their greatest ideas. The Austrian Irish physicist Erwin Schrödinger, who laid much of the groundwork for quantum theory, was among the first to try and understand the messy thermodynamics of the far-from-equilibrium systems that make up the biosphere.

## Schrödinger Solves the Paradox of Life

What Schrödinger noticed and wrote about in his landmark 1944 book, *What Is Life? The Physical Aspect of the Living Cell*, was that instead of drifting toward thermodynamic equilibrium, which for life means death and disorder, biological organisms maintained their living state by consuming and metabolizing energy from the environment. Without a steady supply of incoming energy, equilibrium ensues and life perishes. But by feasting on energy in one form or another, and metabolizing that energy, ordered systems can stay organized without violating the second law.

For example, plants stay alive and avoid decay by absorbing high-energy (low-entropy) sunlight, which is dissipated and returned to the environment as low-energy (high-entropy) infrared light. Some animals stay alive by eating the plants that have internally stored some of that solar energy as sugars and other edible compounds. And some animals stay alive by eating the animals that eat the plants. The energy stored in the molecular bonds of food provides the "chemical fuel" that is used to power all our cellular activities, as well as our thoughts and actions. Almost all the free energy currently being used to support life initially came from the sun, although some organisms feed off minerals in rocks, or the heat energy generated by the earth's core and mantle. As biological systems consume free energy in order to stay far from equilibrium, they simultaneously dissipate it, mostly as heat, producing thermal entropy and exporting it to their surroundings.

The continual growth of biological and technological organization and complexity does not violate the second law of thermodynamics at the level of the universe, because the local reductions in configurational entropy are paid for by the simultaneous increase in thermal entropy caused by life's

constant conversion of free energy into heat. In other words, the cost of building and maintaining structural and functional order at the macro level is the energetic disorder of heat at the micro level. As long as free energy continues to be harnessed and dispersed, the total amount of entropy in the universe increases and the second law remains intact.

What this means is that the universe can grow increasingly organized on a cosmic scale as long as life can find the free energy it needs to build and maintain that order. Luckily for life, the universe offers a vast ocean of exploitable energy to beings that are intelligent enough to know how to extract it. Energy from solar power, nuclear fission, and soon, nuclear fusion will replace the environmentally harmful fossil fuels and be plentiful in supply. Additionally, from Einstein's famous equation $E = mc^2$, we know that every bit of matter in the universe can be converted into usable energy, yielding an amount equal to the mass collected multiplied by the square of the speed of light.

This fact makes it easy to imagine the universe growing increasingly structured and functional as a hyperintelligent civilization spreads through the cosmos at dizzying speeds, transforming all the energy and matter in its midst into exotic forms of biological and computational machinery. But how long can this go on for? Well, that is the million-dollar existential question. The prolific science fiction writer Isaac Asimov called it "The Last Question" in his most critically acclaimed short story about the fate of life in the universe. The real-life version of that dramatic story—the great cosmic battle between life and entropy, order and decay—is reserved for Part Three. As a bit of a spoiler, I will say that I think Asimov would be happily surprised with the likely outcome, which he did not have the conceptual tools to understand back in 1956 when he wrote his visionary tale.

The concepts covered in this chapter provide the foundation of the new cosmic narrative that is the basis of this book. They may take some time to understand on an intuitive level, but as the narrative continues to unfold, their significance will become increasingly clear. We have explained how life thrives and why neither its existence nor its expansion violate the second law of thermodynamics. Biological life reduces configurational entropy and structural disorder by increasing thermal entropy and energetic

disorder. But what has not yet been explained, and what Schrödinger could not address, is how or why biological systems emerge from random molecular chaos in the first place.

Quite surprisingly, new research suggests that life not only abides by the second law of thermodynamics but also emerges *because of it*. The next chapter in our cosmic journey explains how, as a result of entropic forces, nature builds biological machinery. What then is the thermodynamic purpose of biology? To open new channels for planetary energy flow in an effort to more efficiently dissipate gradients and achieve thermodynamic stability.

# 3

# Unraveling the Mystery of Life

As we continue on our journey, let us quickly be reminded of the new cosmic narrative that is being proposed, which posits a self-organizing cosmos. As the universe evolves through time, it tends to move toward an increasingly ordered, complex, and functional state. This cosmic developmental process enables the emergence and evolution of life, which in turn enables the process to continue *indefinitely*, or until maximal complexity and computational power is achieved. Explaining exactly how and why this must occur as a matter of physical and logical necessity is the mission of this book.

This version of cosmic evolution at first seems impossible, or at least improbable, because the second law of thermodynamics says the total amount of entropy in the universe must always increase as time moves forward. But as we learned in the previous chapter, entropy exists in various forms, such that the universe can grow more ordered if free energy is converted into waste heat in the process. More specifically, *configurational entropy* ("disorder") can be reduced and suppressed locally by life as long as *thermal entropy* is continuously produced and exported into the surroundings.

## Extending Thermodynamic Theory

Schrödinger was among the first to popularly explain that this is what life on Earth is doing, revealing biology to be a fundamentally thermodynamic

phenomenon. As open systems, which exchange energy and matter with their surroundings, living organisms can evade the tendency toward disorder and thermodynamic equilibrium—in other words, death and decay—by extracting the free energy that is all around them. For plants, the energy source is sunlight. For animals, the energy source is food, which can be plants or other animals.

Although Schrödinger explained why life does not violate the second law, he didn't attempt to explain how it first emerged, how far it could spread, or the likelihood of such an event happening elsewhere. Despite the progress made in understanding the thermodynamic nature of biological systems in the first half of the twentieth century, life's origin remained a mystery, and Boltzmann-inspired logic perpetuated the assumption that abiogenesis was a fluke occurrence with a very small probability of happening anywhere else.

This depiction of life's emergence as highly improbable was responsible for a number of dubious philosophical conclusions. It fueled creationists' belief that life's existence was proof of divine intervention and led secularists to believe that biology was something of a cosmic quirk: a small and transient pocket of order in an overwhelmingly hostile and lifeless universe. Under the standard secular scientific view, which was the reductionist worldview, there was no reason to think that life would outlast the sun that powers it, and certainly no reason to believe that it could someday acquire the power to influence the large-scale evolution of the cosmos.

However, in the 1960s, a trailblazing Belgian biochemist with a different cosmic vision rose to prominence. According to Ilya Prigogine, we do not inhabit a world that is growing more disorganized and breaking down. Instead, we find ourselves in a universe that is far from thermodynamic equilibrium, with a vast abundance of energy sources and gradients that cause energy to flow in and out of countless open systems. In response to this incoming flow, patterns, organization, and functionality emerge naturally and predictably. Nature is not only destructive but also inherently and wonderfully creative.

# Spontaneous Self-Organization and the Second Law

Want to see an example of natural and spontaneous order? Simply fill a tub with water and then remove the plug. Instantly and entirely reliably, a vortex will emerge over the drain. The water molecules in the vortex do not move about randomly like the ideal gas molecules in the container envisioned by Boltzmann. Rather than behaving entirely chaotically, they move coherently and form an ordered dynamical structure that is steady, stable, and elegant in design. These kinds of structures are common in nature, from the whirlpools that form over sinkholes in lakes to the wind cyclones created by hurricanes and tornados. While in our everyday experience these formations are short-lived, they need not be. The whirlpool over the drain in your tub will remain as long as you keep the water running, and the famous Great Red Spot on Jupiter is a gigantic hurricane that has been swirling around for hundreds of years.

Observing the frequent emergence and persistence of dynamical order in the natural world, Prigogine knew that classical thermodynamic theory needed to be radically extended. The equations of Boltzmann's statistical mechanics are accurate only for isolated systems at or near equilibrium. At equilibrium, there is nothing but random motion and overall stagnation. But systems that are far from equilibrium—which are pretty much the rule and not the exception here on Earth—behave quite differently and, therefore, require new mathematics and foundational principles. Schrödinger said that understanding the behavior of life would reveal "new laws of physics," and these new laws would radically change the way we think about the second law of thermodynamics.

The second law is equally important in equilibrium and far-from-equilibrium conditions, but as we will see, its consequences are reversed in the latter. Rather than dissolving, order emerges and persists. This has to do with the intimate connection between entropy production and the spontaneous pattern formation, which Prigogine recognized was of great importance. Structures like cyclones and vortices don't emerge in spite of the universe's tendency to become more entropic; they form because of it.

Recall from the last chapter that the second law is an expression of nature's intolerance of gradients. This is the reason why heat spontaneously flows from a hotter to colder location until the temperature difference, or the thermal gradient, is eliminated. When a large enough difference in temperature, pressure, chemical concentration, or charge exists between a system and its surroundings, flow will spontaneously occur until the difference is reduced and equilibrium is achieved. This balancing flow dissipates energy gradients, producing entropy—energy that is not available for work.

Now here is where things get interesting. Since nature loathes gradients, the effort put forth into reducing them is proportional to the magnitude of the gradient. When gradients become large enough and sufficient free energy becomes available, a cycling structure will *spontaneously emerge* to enable the gradient to dissipate more efficiently. Because the universe "wants" to produce entropy at the fastest possible rate, it will actually create organized systems that help it achieve its thermodynamic goal more efficiently.

To the conscious observer, these emergent structures seem to materialize out of thin air, built by the "invisible hand" of nature. Take a hurricane for example. The cyclone of wind forms when the temperature difference between a warm ocean and a cold atmosphere becomes great enough to trigger spontaneous emergence. The spiraling winds that form serve to dissipate the gradient by cycling the air, mixing the warmer and cooler molecules until a balanced state of thermodynamic equilibrium is achieved. The dynamic rotating structure persists until the gradient is eliminated and energy flow ceases.

Spontaneous emergence is described as *self-organization* because the emergent structure arises due to the laws of nature and the collective interactions of the system's components, and not because of the actions of an external agent, like a human engineer or a supernatural deity. It is worth noting that it may be a bit misleading to call the order-forming process "spontaneous," as it is always driven by a flow of energy. For this reason, self-organization occurs only with open systems in far-from-equilibrium conditions.

Prigogine named these spontaneously self-organizing systems *dissipative structures* to emphasize their dependence on energy and their

thermodynamic function, which is to minimize free energy and produce entropy. Dissipative structures reveal that entropy production can be as much of a motor for order and complexity as it is for disorder. This discovery fundamentally changes the way we think about the second law: it is both the destroyer and creator of organized structure in nature. "Non-equilibrium brings 'order out of chaos,'" is how Prigogine liked to put it.

Fortunately for Prigogine, dissipative structures are easy to create in a controlled lab setting. Hit a many-particle system with a steady flux of energy, like a stream of heat or an electrical current, and watch it self-organize like magic. As the energy flows through the system, it drives it away from the disordered equilibrium state toward a "steady state" of dynamical organization and function.

This effect is easily achieved by simply heating some water in a pan on the stove. As the bottom layer of the liquid heats up, a temperature difference between it and the cooler surface layer above begins to grow. The temperature gradient causes heat to flow from the bottom layer toward the top layer, creating an upward current. As the hot (excited) water molecules move upward, some have enough energy to escape as vapor. Evaporation cools the molecules at the surface, causing them to sink back toward the bottom. These heat-induced currents create patterned flows across paths of least resistance that form organized hexagonal structures called *convection cells*, also known as Bénard cells.

A more complex dissipative structure can be thermodynamically conjured by creating a chemical concentration gradient that drives a chemical system away from equilibrium. This happens in what's known as the *Belousov-Zhabotinsky reaction*, which Prigogine studied and modeled mathematically. In this example of emergence, the appropriate mix of chemicals produces a reaction solution with striking spatial and temporal patterns visible to the naked eye. Not only that, the color of the solution oscillates from blue to red at regular time intervals, creating a bona fide chemical clock! This colorful clock is a stunning example of how a thermodynamic imbalance brings ordered machinery into existence.

## Attractors Create Order by Constraining Chaos

Dissipative structures come in many different shapes and sizes, but they all cycle around stable dynamical states of order that complexity scientists call *attractors*. An attractor is a set of states or configurations that a dynamical system naturally tends to evolve toward, regardless of its "initial conditions"—how the system started off. Think of a ball rolling down a hill into a valley. The basin of the valley is an attractor. It doesn't matter where on the hill the ball started off, it's eventually going to end up at the bottom. And because being at the bottom represents a state where the system has minimized its potential energy, once the ball gets there, it typically stays there. Of course, an injection of energy could move it out of the attractor, but escaping the attractor does not happen spontaneously.

While an equilibrium state of complete chaos and maximum disorder is an attractor for isolated systems, open systems being pushed by a flow of energy inevitably move toward *non-equilibrium attractors*—steady states of order that manage to persist by continually absorbing and dissipating energy from the environment. While an isolated system of interacting molecules at thermodynamic equilibrium explores its *configuration space*—the *three-dimensional space of possible configurations*—in an essentially random manner, an open system driven by a steady flow of energy will evolve deterministically toward an attractor that represents the region of the configuration space that corresponds to a stable molecular arrangement.

Since a stable state of non-equilibrium persistence requires continual free energy to be maintained, the attractor that the system evolves toward is configured for optimal energy absorption and dissipation. In other words, a natural pressure to minimize free energy and produce entropy spontaneously creates an ordered structure with an objective function. With the emergence of dissipative structures, we see the emergence of purpose in nature, and that cosmic purpose is associated with maximizing the rate of entropy production.

If the dissipative system is driven even further away from equilibrium by an increase in the rate of energy flow, and if it does not crack under pressure, another critical threshold will eventually be crossed, and the evolving system will enter a new attractor—one that corresponds to a configuration that further enhances energy absorption and dissipation. This transition to a state of higher order is known as a *phase transition*, and although we can roughly predict when such a transition is likely to occur,

predicting the properties of the new phase is often difficult if not impossible, as the process is as chaotic as it is orderly. This means that as long as energy is flowing throughout the cosmos, nature will always present us with interesting surprises, designed by the delicate dance between order and chaos that produces emergence.

## The Birth of Non-Equilibrium Thermodynamics

Prigogine's experimental and theoretical work on dissipative structures would win him a Nobel Prize in 1977, but more importantly, it would give rise to entirely new fields of physics and chemistry, like *non-equilibrium thermodynamics* and the closely related *non-equilibrium statistical mechanics* (often lumped together as *non-equilibrium physics*). This rapidly growing area of research has become a pillar of complexity science and is significantly influencing biology and many other fields, due in no small part to the scientists at the Santa Fe Institute, a theoretical research center established in 1984, and their collaborators at universities and think tanks around the world.

While Prigogine's equations described only the emergence and behavior of relatively basic physical and chemical dissipative structures—certainly nothing as complicated as the formation or operation of a cell—the grander implication was obvious: living organisms, which maintain their ordered, far-from-equilibrium state by capturing and degrading free energy through the chemical process of metabolism, are highly complex dissipative structures. If you cut off their energy supply, like all dissipative structures, they will break down and disintegrate into chaos as the system relaxes to thermodynamic equilibrium.

If living organisms are dissipative structures, Prigogine reasoned, then maybe life's origin isn't so hard to explain after all. The first form of life could have emerged in response to far-from-equilibrium conditions, forced into existence by a natural pressure to dissipate a thermal or chemical gradient (or some combination of them). In such a scenario, abiogenesis is reconceptualized as a thermodynamic event that opened up new energy flow channels on Earth to facilitate entropy production. And as these channels emerged and expanded to relieve thermodynamic pressure, the

increased energy flow stimulated further chemical and biological self-organization, eventually giving rise to the interconnected global living network that we call the biosphere. This was clearly the hypothesis that Prigogine had in mind when he wrote, "At equilibrium matter is blind; far from equilibrium it begins to 'see.'"[1]

## A New Paradigm for Understanding Life

Non-equilibrium statistical mechanics provided a paradigm that finally allowed scientists to begin to describe mathematically both the dynamics and thermodynamics of the complex chemical whirlpools that we call organisms. As this new paradigm continues to grow in popularity, life is gradually losing its reputation as a vastly improbable phenomenon. It seems that with the appropriate far-from-equilibrium conditions, such as those present on the early Earth, the biological state of "being alive" is an attractor whose emergence is statistically likely, if not downright inevitable, given enough time.

Prigogine's story sounded great, but it was not without its problems. Despite his compelling theoretical vision, he could not demonstrate his model of abiogenesis in the lab, nor could others, and this wasn't from a lack of trying. Not only that, his theory of dissipative structures lacked a clear mechanism for molecular self-organization that could be described by a precise mathematical equation. However, in the new millennium, significant progress would be made in both of these domains.

Prigogine's seminal work inspired many scientists to pursue the origin-of-life problem from a thermodynamic and statistical perspective, yielding thousands of papers published in top-tier journals over many decades. Building on this research and more recent work by chemist Gavin Crooks at the Lawrence Berkeley National Laboratory, a young physicist at the Massachusetts Institute of Technology (MIT), Jeremy England, proposed a Prigoginian theory called *dissipative adaptation*, which would capture the attention of mainstream media.

The basic theory, first outlined in detail in a 2015 paper published in *Nature Nanotechnology*, introduced a formal mechanism for energy-driven

molecular self-organization. But the real excitement came in 2017, when England and his colleagues demonstrated the mechanism with a clever use of modern technology. Computer simulations allowed them to digitally construct "toy" chemical networks that could be visually observed as they were incrementally driven away from equilibrium.

What England and colleagues saw was astounding. In response to a simulated flow of thermal energy or an induced chemical gradient, randomly distributed "molecules" began gradually aligning themselves into ordered collective arrangements that would be considered extremely improbable from the classic thermodynamic perspective of Boltzmann. More specifically, the interacting chemical units self-assembled into cohesive configurations that allowed the network of particles to more efficiently absorb and dissipate the free energy flowing into the system. In other words, driving the system away from equilibrium caused a rudimentary dissipative structure to not just emerge but also *evolve*.

England described this process as "adaptation" because the configurations that the emergent molecular system evolved toward were specifically those that better resonated with the frequency of the energy source that was doing the driving. You could say those configurations were more "fit" for survival, as they better absorbed the free energy that kept the system in existence. The results were clear: given the appropriate far-from-equilibrium conditions, a randomly oriented system of molecules will reliably evolve toward a configuration that is more complex and fine-tuned to the surrounding energy landscape. Suddenly, it seemed that Darwin's theory had something to say about the origin of life after all, but it was clear that it would have to be expanded to include dissipative evolution and other nongenetic forms of adaptation.

While England's abstract digital demonstration may not perfectly reflect how actual chemical networks self-organize—simulations are always gross simplifications of messy real-world systems—recent physical experiments carried out independently by other research groups have demonstrated the same fundamental process, and many of these studies were published prior to England's theory. For example, in a 2013 experiment published in the online journal *Scientific Reports*, a group of researchers from Japan showed that simply shining a light on a cluster of silver nanoparticles caused them

to self-assemble into a more ordered collective structure, with the ability to capture and dissipate more energy from the light source.

More extraordinary examples followed, like the 2015 study published in the journal *Physical Review E*, which demonstrated dissipative adaptation at the macroscopic level. When conduction beads were placed in oil and struck with steady voltage from an electrode, the beads collectively formed an intricate dissipative structure with "wormlike motion" that persisted as long as energy flowed through the system. Not only that, the emergent structure avoided configurations that lowered available energy and moved toward positions that did the opposite, which the authors labeled "energy-seeking behavior."

This suggests that dissipative adaptation could be the mechanism that brought about the emergence of what we think of as agency, intent, and purpose—an idea that we will explore in detail in Chapter Six, which explains how the thermodynamic process England identified might create information and enable information processing, or *computation*.

Though dissipative adaptation occurs long before a system has genetic material or the ability to self-replicate and mutate—a requirement for Darwinian evolution of the kind to which we are accustomed—the basic chemical system does undergo a type of thermodynamic evolution that involves blind variation and natural selection. As the ensemble of molecules is driven away from equilibrium, its collective form rapidly moves through a number of different structural and functional arrangements. Out of these many variants, only the configurations that allow the emergent dissipative system to extract the energy it needs to persist get to "stick around," while all the others are passed over and "forgotten" by the system. Moreover, because configurations that dissipate more energy produce more entropy, they tend to be more irreversible due to the second law of thermodynamics and the one-way arrow of time it dictates.

In a nutshell, the second law favors the emergence and persistence of increasingly complex and functional energy-extraction machines, and when those machines become sufficiently complex, we call them organisms. A 2017 paper coauthored by physicist Susanne Still calculates that the odds of life emerging on a planet increase significantly as the system is driven further away from thermodynamic equilibrium.[2] This suggests that for

planets with similar thermodynamics and geochemistry as the early Earth, life is essentially guaranteed to emerge and evolve. That means we can no longer think of life as a cosmic accident.

Dissipation-driven self-organization can look like magic, but it requires no supernatural force to proceed. It is thermodynamic in origin, Darwinian in nature, and mathematically describable. The order-creating power of energy flows might sound like something out of a New Age bible with a mystical and pseudoscientific origin story, but that doesn't change the fact that it really is how nature works.

## Collective Autocatalysis Catalyzes Creation

While England's simulation studies and the analogous physical experiments tangibly demonstrate how a clump of molecules can self-assemble into an ordered structure with superficially lifelike behavior, dissipative adaptation only takes us so far. Even the simplest conceivable cell contains many thousands of complex molecules that all participate in an elaborate and highly coordinated chemical dance. So, the question is: How does a small chemical system get amplified into a vast biochemical network that is complicated enough to perform metabolism and self-replication? What the system needs is a way to quickly ramp up its size and chemical complexity, and this is where the dissipative process known as *autocatalysis* enters the picture.

We can think of autocatalysis as a process that allows a chemical system to quickly self-repair or self-amplify in size and complexity through a series of linked reactions that establish a positive feedback loop. While this process sounds complicated, and it is, we can imagine its simplest version to illustrate how it might work in theory. Picture a "chemical soup" in a flask that has an abundance of a few simple molecules. Through random interactions, or some mechanism like dissipative adaptation, the small molecule A reacts with small molecule B to form a larger molecule C. But C happens to be special because it is a *catalyst*—its presence naturally speeds up the reaction between molecules of type A and B, which in turn rapidly produces more of C. In this scenario, the production of a single molecule C, which will occur at some point simply due to molecular

fluctuations, triggers a fast and furious cycle that makes more and more of the complex molecule from its simple precursors. This is autocatalysis.

But how exactly might this take us from a simple chemical network to something as complicated as an actual cell? Well, you could imagine a complex array of various molecules carrying out *collective autocatalysis*, thereby generating an increasingly large network of reaction cycles that are interlinked and self-reinforcing. Paul Davies, an influential theoretical physicist and author who has recently directed his efforts to the origin-of-life problem, explains this beefed-up version of autocatalysis nicely in his 2018 book *The Demon in the Machine*:

> There is thus a feedback loop: groups of molecules catalyze their own production. Scaling this up, there could be a vast network of organic molecules forming a quasi-stable system of autocatalysis, with many interlocking feedback loops, combining in a tangled web of reactions that is self-sustaining and robust.

By "self-sustaining," Davies means that an autocatalytic system can continually create and repair all of its chemical components, as long as its environment supplies the energy needed to drive the cycle. When a system can maintain and renew itself, it has moved one step closer to being "alive." In the philosophical literature, this remarkable property of life is known as *autopoiesis*, and we will come back to this concept in Part Two.

Collective autocatalysis isn't just a neat theoretical concept. It has been demonstrated in the lab using simple precursor molecules. Scientists have not been able to make something as complicated as a cell artificially, but they're inching closer to that goal. In 2004, a team at the Scripps Research Institute in La Jolla, California, created an autocatalytic set of nine peptides from amino acid precursors, bringing us one step closer to creating proteins, a necessary ingredient for life.

In 2012, scientists at Portland State University in Oregon were able to produce an RNA-based autocatalytic set capable of self-replication from non-replicating RNA fragments. To be clear, these sets are neither adaptive nor autonomous enough to deserve the label of "living," but they demonstrate how chemical systems that use positive feedback can grow increasingly complex through self-amplifying cycles. They also suggest that

the formation of a biological cell was an incremental process rather than a singular event.

While the notion of autocatalytic sets has been around at least since the late 1960s, first described by scientists like Manfred Eigen, Prigogine, and Stuart Kauffman, dissipative adaptation makes it much easier to explain how they might get going. Without such a mechanism, the initial formation of the complex autocatalytic set that led to life would seem to require a highly unlikely coming together of just the right molecules in just the right way. But if collective autocatalysis is facilitated by dissipative adaptation, abiogenesis may have been the inevitable outcome of persistent driving forces created by far-from-equilibrium conditions. In this case, life's self-assembly is the product of an organizing law rather than the product of chance assembly. Your single-celled ancestor was not a fluke of nature!

Despite the mechanistic explanations for how life can naturally emerge from blind physical processes, we are left with our previous problem. Autocatalytic sets, as complex as they may be, are still a far cry away from cells, which contain a labyrinth of nucleic acids and proteins intricately bundled inside a fatty membrane composed of lipids. And when a chemical mixture of the basic elements that make up life—carbon, hydrogen, oxygen, nitrogen, sulfur, and phosphorus—is driven away from equilibrium by heating it or passing an electrical current through it, we typically don't get any of those large organic molecules.

Only relatively simple amino acids or peptides can be produced, and although this created a lot of excitement in the 1950s when it was first demonstrated by biochemists Stanley Miller and Harold Urey, enthusiasm waned when scientists realized that such simple methods were unlikely to yield anything much more complex. The abject failure to re-create life in the lab suggested that its emergence could not have been the simple result of a warm pond getting hit by jolts of lightening, or even strong and steady sunlight, as the popular "primordial soup" theory proposed.

So, the lingering questions are, what were the particular far-from-equilibrium conditions on Earth that created an actual cell, how exactly did it happen, and what makes it so difficult to replicate now? While we may never have all the details of how it happened on our planet, origin-of-life researchers and complexity scientists have recently sketched out a

compelling explanation that is consistent with the thermodynamic narrative introduced in this chapter.

# 4

# The Emergence of Life on Earth

In the prior chapter, we identified precise thermodynamic mechanisms that explain how inanimate molecules can naturally self-assemble to form complex chemical networks. However, these processes don't seem to get us all the way to something worthy of being called "alive," and attempts to manufacture a cell have ultimately been unsuccessful. Given the Herculean challenge posed by the task, we should not be too surprised by our failure.

To be able to self-replicate, a basic necessity of life, a cell requires a variety of complex biomolecules acting in coordination. The most important of these molecules, DNA and RNA, store and transmit genetic information. They provide the blueprints and instructions for building the cell. Proteins are needed to physically construct the cell according to those instructions. Also, special proteins called *enzymes* are needed to catalyze all the chemical reactions a cell needs to function and metabolize energy. None of these molecules have been created in the lab from the basic atoms that comprise them, and it has been hard to imagine how any of them could arise on their own spontaneously, since the creation of DNA or RNA requires proteins and the creation of proteins requires DNA or RNA.

Thus, we have a mystery similar to the old conundrum that asks, "Which came first, the chicken or the egg?" On top of that, these large molecules need to be bound together and protected from the environment by a semi-permeable membrane composed of fat molecules called *lipids*. However, the origins of all these critical complex biomolecules—DNA, RNA, proteins, and lipids—become a lot less mysterious if we consider the

process of metabolism as the starting point. Although we may never know the exact details of how abiogenesis occurred on Earth, we can at least construct a theory that best explains the evidence with the least number of assumptions.

## Energy Flow, Organization, and Metabolism

Prigogine wasn't the only scientist of his generation thinking about the origin of life from the perspective of non-equilibrium thermodynamics. In a somewhat obscure but influential book called *Energy Flow in Biology*, published in 1968, Yale University biochemist Harold Morowitz wrote a memorable and prescient statement: "The energy that flows through a system acts to organize that system." Like Prigogine, Morowitz believed that life's essential feature was the process of metabolism, as it is what sustains the organism's structure and function by allowing free energy—extracted from sunlight or food—to continuously flow through the system, revitalizing it. In his mind, explaining the origin of life begins with explaining the origin of metabolism.

Metabolism is made possible by sets of chemical reactions occurring inside cells, catalyzed by certain enzymes. Some metabolic pathways break down large organic compounds to generate energy, while others use energy to build up compounds from smaller molecules, permitting organisms to grow and repair themselves. In high school biology, we are taught that a chemical reaction loop called the *Krebs cycle*, also known as the *citric acid* or *tricarboxylic acid (TCA) cycle*, allows humans to turn food into usable energy with a little help from the oxygen we breathe. But what does this all have to do with the origin of life?

The first living organism, a single cell, required some primitive form of metabolism to survive, but it couldn't get energy from sunlight, since photosynthetic machinery is relatively complex and presumably also came later, via biological evolution. Also, this protocell (a simplified version of a cell that is still capable of growing and replicating) couldn't get energy by eating other organisms, since there were none around. How the first life-form got its energy was a great mystery until a new piece of the puzzle emerged in 1977, when ecosystems filled with new exotic life-forms were

discovered deep in the ocean in conditions previously thought to be too extreme to support biology. Miles below the water's surface on the ocean floor, these organisms were clustered around majestic geological structures known as hydrothermal vents.

These geysers, or hot springs, are underwater volcanoes that transport heat up from the earth's smoldering interior through convection. There are many such vents around the world, existing at mid-ocean ridges formed by the spreading apart of tectonic plates. The simplest organisms living around these vents are a class of bacteria called *reductive autotrophs*, which feed on the reservoirs of geothermal and geochemical energy produced by the mixing of hot magma, seawater, and rock minerals. They, therefore, require neither sunlight nor other organisms to sustain themselves. Scientists learned that these microbes carried out metabolism by running a stripped-down version of the familiar Krebs cycle, but in reverse. By *running it backward*, the chemical reactions driven by a relatively simple autocatalytic set could build and repair the organism's biological structure using only the small inorganic molecules that are provided by the geochemistry in hydrothermal vents.

This Krebs cycle in reverse was named the *reductive TCA cycle*, and soon biochemists and molecular biologists would recognize its fundamental importance. In its most primitive form, the reductive TCA cycle doesn't need an entire cell to proceed. A chemical network composed of as little as eight carboxylic acids, simple organic molecules with only a few carbon atoms, could drive the cycle and synthesize biomolecules of arbitrary complexity solely from inorganic inputs. In fact, scientists showed that the cycle could produce all of the basic organic compounds that make up every known living organism—including each protein, nucleic acid, and lipid—using only carbon dioxide and hydrogen.

It wasn't long before some imaginative origin-of-life researchers—perhaps most prominently Michael Russell at the California Institute of Technology—saw a possible connection between the geochemistry at hydrothermal vents and the metabolic cycles that sustain life. The hot and chemically diverse environment would provide not only the molecular building blocks, such as carbon, hydrogen, oxygen, nitrogen, and phosphorus, required for abiogenesis, but also the high temperatures and

pressures needed to organize them into a complex reaction network like the autocatalytic set that runs the reductive TCA cycle.

This wasn't all hydrothermal vents could provide. In addition to the necessary precursor molecules and the energy flows needed to assemble them, their rocky structures contained countless pores that could function as compartments where autocatalytic sets might form and be shielded from disruptive chemicals and external perturbations. In other words, collective autocatalysis capable of building biological machinery would not require a lipid membrane to get started. The pore could serve as a *Markov blanket*—a barrier that preserves an ordered system and separates it from its environment.

Nor would the protocell need proteins to catalyze its reactions, since the rocks provided mineral compounds and other small molecules that could act as catalysts, even if they did the job poorly by comparison to today's enzymes. With such possibilities, it is not difficult to imagine how over time an autocatalytic set running the reverse Krebs cycle could generate all the proteins, nucleic acids, and fats necessary for cellular structures and functions, such as the more sophisticated modes of metabolism, self-regeneration, and eventually, self-replication.

Decades after the publication of *Energy Flow in Biology*, starting around the turn of the twenty-first century, Morowitz teamed up with physicist Eric Smith at the Santa Fe Institute to publish a series of papers that outlined their version of this new origin theory, based on non-equilibrium thermodynamics and the work of dozens of like-minded researchers. According to this new narrative, life on Earth emerged as a self-organized dissipative system, driven into existence by the thermal and chemical gradients produced in hydrothermal vents, or some geologically similar location where rock, water, and intense heat intersect.

## Life: The Universe's Great Balancing Act

On early Earth, which was undergoing rapid structural and chemical change, the kind of activity we now see at vents would have been much more widespread. This presumably produced huge amounts of potential chemical energy that accumulated into vast reservoirs. With no life-forms to

graze on that energy, untapped gradients created a planetary-scale energy imbalance that caused global thermodynamic instability. As a result of this tension, autocatalytic chemical sets resembling primitive metabolism spontaneously emerged to unlock and dissipate free energy excesses at locations where stress was greatest. Since the geological activity was constantly creating chemical energy, the metabolic dissipative structures persisted, giving them a chance to evolve and develop.

Presumably, as the complexity of the reaction networks inevitably increased, an autocatalytic set running something like the reductive TCA cycle synthesized the essential cellular components and gradually incorporated them back into the set that spawned them. As this process proceeded, the organic chemical network grew increasingly complex, and a self-renewing system (*autopoietic* agent) was transformed into a self-replicating system (organism).

This explanation not only accounts for life's origin but also the formation and maintenance of the biosphere as a whole. In Smith's words, the interconnected network of living organisms that covers Earth's surface is a "relaxation channel" for relieving planetary energy stresses, similar in function to a bolt of lightning, another dissipative structure.[1] Lightning spontaneously forms when the electrical charge separation between the ground and upper atmosphere becomes sufficiently large, collapsing the gradient by creating a flow channel. Similarly, life provides a flow channel by opening up a whole new domain of chemical reactions—known as organic chemistry—that can better dissipate accumulated geochemical energy.

According to the vision of Smith and Morowitz, the "collapse to life" was as inevitable and natural as water flowing downhill to find its lowest energy state. And just as flowing water carves a channel into the hillside that becomes progressively deeper with time, resulting in greater flux, energy flows reinforced and strengthened the chemical channels that we call *metabolic pathways*. As these pathways developed, they provided a solid foundation for genetic and enzymatic complexity to be built upon. The theory put forth by Smith and Morowitz forces us to see biology in a whole new light. Life is not just some molecular fluke. It is a lawful planetary process that provides a solution to a chemically unstable abiotic state. In the

words of their colleague at Arizona State University, Sara Walker, "Life is not something that happens *on* a planet; it is something that happens *to* a planet."[2]

This thermodynamic paradigm that explains abiogenesis may also help explain biological evolution and ecosystem development. Since the geochemical processes causing life to emerge continually produce chemical energy, and because solar energy is constantly striking the surface of the planet from above, biology doesn't just come and go like a lightning bolt or a tornado. Instead, it persists and propagates as it feeds on the fuel, the free energy that keeps it organized and running. As the energy flows drive dissipative and biological evolution over long spans of time, life can't help but discover ways to tap into new sources of energy, like sunlight, oxygen, flesh, and fire.[3] As more energy cycles through the living network, there is more energy to power self-replication, and organisms make more organisms, which make more organisms, which diversify and link up to form complex food webs.

In this way, ecosystems can be seen as *high-level autocatalytic sets*, with organisms and species carrying out collective metabolism in the same loopy way that chemically linked biomolecules do in cells. As autocatalytic webs of creatures "renew their components" through biological reproduction with mutation, life on Earth ceaselessly regenerates, diversifies, and complexifies through an autopoietic process, that is, a self-renewing process. We will come back to this idea in Part Two when we discuss the evolution of the biosphere, but it is worth noting now that the story of life—and the story of the cosmos itself—is a tale of *loops and levels*.

From this new account of abiogenesis, we can see that the transition from nonlife to life doesn't happen in an instant. Instead, it occurs in a series of steps, as metabolism builds up functional architecture little by little, layer upon layer. In the epic monograph *The Origin and Nature of Life on Earth*, published in 2016 by Cambridge University Press, Smith and Morowitz describe this stepwise developmental process as a "cascade of phase transitions."

# Phase Transition Theory and the Origin of Life

A *phase transition* refers to a sudden holistic change in the overall arrangement of a system's structure, and in turn, its function. Most people are familiar with basic phase transitions, taught in high school chemistry, such as liquid water freezing into ice under cold temperatures. During this phase change, the $H_2O$ molecules are transformed into a more stable and orderly configuration, from a liquid phase to a solid one. More complicated phase transitions occur in nature when a group of interacting components self-organize into a coordinated and stable collective entity. This happens at varying levels of abstraction. An easy-to-visualize example is when flocks of birds fly in unison, forming shape-shifting geometric patterns that seem to constrain and guide the pack, a phenomenon that is stunning to watch. We can also think of the emergence of human civilization as a great phase transition, whereby behavioral chaos collapsed into social order. Non-equilibrium phase transitions are critical to the story of cosmic evolution, since they provide the mechanism through which new levels of complexity and organization spontaneously come into existence.

How a complex system suddenly transitions into a more organized state is not entirely understood, but it can only occur when the system's components are collectively interacting in such a way that the activity of its parts becomes increasingly coordinated and statistically correlated. This does not mean that every component needs to interact with all of its neighbors, but they should all be linked together in a way that allows the effects of local interactions to propagate through the entire network like a chain reaction. When the collective effects of these mutually reinforcing interactions reach a critical threshold, in a manner that is purely mechanical but also quite mystical in appearance, global patterns of synchronized activity suddenly emerge, and the system gains some new property or function, appropriately called an *emergent property*.

To add further mystique to the phenomenon of emergence, it is difficult if not impossible to precisely predict when a transition will occur using known mathematics, as the phase-shift process involves "chaotic" self-amplifying cycles that simply can't be computed or modeled with precision. This is because of the *butterfly effect*, which refers to how a very small and

seemingly insignificant action, like the flapping of a butterfly's wing, could get violently amplified through chaotic cycles such that it causes a tornado on the other side of the world. What this effect means is that if you try to model a physical system that has chaotic dynamics, the measurement error that can be expected from any conceivable instrument, as tiny as it may be, will be enough to throw off any precise calculation of a future state.

For this reason, the behavior of life, from honeybees to humans, is believed to be unpredictable, not just in practice but in theory. Not even the world's most powerful computer can predict the precise behavior of life. However, far-from-equilibrium phase transitions can be predicted in a probabilistic sense using the emerging mathematical tools of non-equilibrium statistical mechanics, so scientists are far from hopeless in the mission to understand and predict reality, including the collective behavior of agents.

Because chaotic feedback loops can quickly amplify the size of a chemical system and its metabolic complexity, an autocatalytic network can undergo a series of phase transitions in rapid succession if the system is being driven by a strong enough energy gradient. Driven by persistent energy flow, the complexity of a dissipative chemical network steadily increases, new features are forged, and finally, self-organization spawns a cell.

Like Jeremy England, Eric Smith and Harold Morowitz attribute abiogenesis to a Darwinian-like process that involves natural selection acting on thermodynamic variation. If a phase transition brings about a collective configuration that improves the system's ability to absorb the free energy that sustains it, that ordered phase will endure relative to others. Since phases that do not enhance energy utilization and dissipation will dissolve about as quickly as they emerge, an autocatalytic system undergoing amplification will keep reverting back to the previous stable configuration until it discovers a superior (more stable) state. And because collective molecular arrangements that can extract accessible energy tend to be more complex, progress toward increasingly sophisticated dissipative structures is not improbable—in far-from-equilibrium conditions it may be unavoidable.

# Questions and Conclusions

This perspective helps explain why once life formed, it was able to survive in a harsh and constantly changing environment. As the first complex adaptive system evolved through successive irreversible transitions, it became incrementally more robust, ordered, and compartmentalized. The multiple phase-transition model of biological emergence elegantly explains life as an expected consequence of a natural evolutionary process rather than the result of a divine or statistical miracle. With our new thermodynamic theory of life, we can answer some outstanding existential questions:

*Was life on Earth an anomalous accident?* No, biology was the inevitable result of dissipative adaptation, autocatalysis, and non-equilibrium phase transitions. These organizing mechanisms built a biochemical flow channel to resolve a thermodynamic imbalance arising from free energy buildup on a sunny, wet, geochemically active planet.

The statistical inevitability of both the organism and the biosphere is not a result of magic, but the non-equilibrium dynamics that force these attractors into existence can certainly look magical. We can say with semantic justification that energy-driven self-organization is the magic of nature and natural law, *true* magic, and its mystical aesthetic arises from the human perception of unavoidable complexification over time and the once unthinkable cosmic destiny implied by such a process should it turn out to be fundamental and continual. The only thing predictable or determined is the general attractor, not the route there. A rock atop a hill will necessarily roll down the slope, but the precise path it takes is largely unpredictable. A biosphere will tend to produce an increasingly intelligent species, but the route to general intelligence likely follows a different evolutionary path on each life-producing planet.

*Is there life on other planets?* We should expect life to readily emerge on planets with geochemical conditions that are sufficiently similar to those on Earth around 3.8 billion years ago. The question is: What counts as "sufficiently similar"? Some origin-of-life experts, such as Russell and University College London's Nick Lane, suspect that "life would emerge using the same pathways on any sunny, wet rocky plane."[4] For skeptics'

sake, let's also require such planets to be roughly the size of the earth and a similar distance from a similar-sized star. Astronomers conservatively estimate that trillions of these other "Earths" exist in the observable universe, and billions can be found in the Milky Way galaxy alone. Even if life only emerges on a small fraction of such planets, we are still talking about many millions of planets that harbor life. And, if we want to estimate how long life typically persists on those planets, let us be reminded that on Earth it hasn't blinked out once since it emerged nearly four billion years ago, despite frequent natural catastrophes and constantly changing chemical conditions. For these reasons, most complexity scientists and astrobiologists are optimistic that we will soon detect traces of life on some other planet, perhaps in the next few decades.

*What is the relationship between life and the second law of thermodynamics?* Life emerges in far-from-equilibrium conditions to facilitate energy flow and entropy production. The thermodynamic cost of building and sustaining biological order is paid for by the thermal entropy from the waste heat that organisms constantly pump into the surroundings. In this way, life and the second law can exist in perfect harmony. And they don't just peacefully coexist, the relationship is intimate. Romantically speaking, if entropy could talk to life, we could imagine it would say, "I need you," and life would reply, "You complete me."

*If life was inevitable, how come the cell only emerged once in the history of the earth?* While it is true that abiogenesis doesn't appear to have ever been repeated, it is also true that we haven't found a single place on Earth where life is not present, no matter how cold, hot, deep, or turbulent that place may be. Free energy exists nearly everywhere. Even inside volcanoes and miles below Earth's surface, you are guaranteed to find some organisms grazing on it. This suggests that the collapse to life and subsequent development of the biosphere reduced much of the planet's energy imbalance, eliminating a great deal of the thermodynamic pressure that forced life into existence. If we could wave a wand and magically make all life suddenly disappear, we should not be surprised to see it eventually emerge again, once enough chemical energy has accumulated and a breaking point is reached. If Smith and Morowitz are correct, the spread of life is like a propagating fracture.

*Why has life been so hard to create in the lab?* Life is complicated, and so is geochemistry. Re-creating the precise chemical conditions that existed billions of years ago is unquestionably difficult, as is replicating the extreme temperatures and pressures around hydrothermal vents. Even if you get all that right, the lab only allows small samples of matter to be experimented with. Compare that to the prebiotic Earth, which had miles and miles of highly concentrated areas of geothermal and geochemical activity. Although engineering abiogenesis in the lab may be laborious and costly, there is no reason to believe that it can't or won't be done. Many origin-of-life researchers now believe it may happen soon, possibly within decades. How awesome would it be if extraterrestrial life were discovered around the same time that scientists figured out how to make it from scratch?

By dismantling the narrative that life is an accident or some kind of statistical fluke, these emerging empirical explanations may bring a newfound sense of existential comfort to rationalists. One need not be a card-carrying cosmic optimist to find spiritual satisfaction in the idea that life is the inevitable product of an evolving, self-organizing universe. It is a natural expression of the cosmic code rather than an anomalous accident.

The same set of facts that comfort some may bring others to a different conclusion. In the eyes of the cosmic pessimist, the thermodynamic narrative, as it has been told so far, depicts life as nothing but a glorified dissipative structure and a means of entropy production, no more special or significant than a hurricane or whirlpool and probably just as transient. When the stars burn out and the solar gradients that support complex life disappear, so goes with it consciousness and intelligence. Sentience says "sayonara," and the universe goes back to sleep, this time for good. But it is a mistake to jump to this conclusion. Living organisms are fundamentally different from all the other dissipative structures that emerge in nature because they have *agency*, a property that gives them the ability to control their own fate.

## The Game Changer Is Agency

Agency is difficult to define precisely, but we know it when we see it. While inanimate objects only move when they are pushed by some external force, the movements of systems with agency, called *agents*, appear to be internally generated and geared toward some intrinsic goal. While some dissipative structures, like hurricanes and tornadoes, have chaotic and unpredictable trajectories that might momentarily fool an unsuspecting child into thinking they are alive, soon it becomes apparent that the skittish wanderings are actually random and aimless.

Typical dissipative structures are slaves to the laws of physics and the entropic forces that propel them. Autonomous agents, however, can "push back" on nature and cause things to happen, an ability to affect reality that philosophers call causal power. Defining this term precisely has been a challenge for scientists and philosophers, but we may think of causal power as an agent's ability to initiate new chains of cause and effect.

Because it has no causal power, a nonliving dissipative structure cannot initiate actions or control physical events, making it unable to perpetuate itself by seeking out new energy sources when the supply is running low. On the other hand, if the free energy needed to maintain an organism has run out in its immediate environment, the organism will actively seek it out. This would be like a whirlpool that reaches up and turns on the faucet when it senses that the water supply is about to run out, which of course we never see. In effect, nonliving dissipative structures are doomed to a transient and insignificant existence, inevitably dissolving as soon as the gradients that support them are eliminated and energy flow comes to a halt.

This means a living agent can choose its own path, moving in a manner that is in perfect accordance with physical law, though not entirely submissive to it. A stone must always roll downhill and never up because of the force of gravity, but the same cannot be said for something that is alive. This is true even for simple organisms. For instance, many plants slowly but restlessly move in the direction of the sun, tracking it across the sky each day, a tendency called *heliotropism*. Likewise, single-celled creatures as basic as bacteria swim toward food sources and away from poisons, an ability known as *chemotaxis*. When one energy gradient is tapped, another is actively sought out. Whether such simple systems have any modicum of awareness or subjective experience is a subject of heated debate (and

discussed further ahead in the book). But, sentient or not, these agents move with unmistakable purpose and intent.

Nineteenth-century psychologist William James vividly described this unique characteristic of life with a romantic Shakespeare-inspired metaphor:

> Romeo wants Juliet as the filings want the magnet; and if no obstacles intervene he moves towards her by as straight a line as they. But Romeo and Juliet, if a wall be built between them, do not remain idiotically pressing their faces against its opposite sides like the magnet and the filings . . . Romeo soon finds a circuitous way, by scaling the wall or otherwise, of touching Juliet's lips directly. With the filings the path is fixed; whether it reaches the end depends on accidents. With the lover it is the end which is fixed, the path may be modified indefinitely.[5]

In light of this clear behavioral distinction, a complete explanation of life's origin must answer the philosophically loaded but scientifically valid question: How exactly did molecules acquire agency, control, and purpose? Precisely what process instilled a physical system with goals and the capacity to pursue them? If agency does in fact emerge from thermodynamic mechanisms like dissipative adaptation and phase transitions, then we must explain how and why.

## The Unreasonable Robustness of Biology

Another mystery that is not sufficiently explained by merely labeling life a dissipative structure is the question of why it is so incredibly robust and resilient. Ordinary dissipative structures quickly materialize and dematerialize, typically lasting seconds, minutes, hours, or days. On the other hand, life hasn't gone extinct once since it arose four billion years ago, and we now have some very compelling reasons to believe that it is just getting started.

Most of us have been taught that life is fragile, its existence always at risk, but this is a parochial error stemming from, ironically enough, an

excessively anthropocentric worldview. While this may be true for any single organism or species, it is not true of entire ecosystems—or the biosphere as a whole. We must remember that life is a collective phenomenon. Individuals come and go, but the patterns of living organization that maintain coherent structure and function persist without interruption. The biological network on Earth hasn't just persisted, it has flourished, despite being tested by countless natural disasters and multiple periods of mass extinction. After each catastrophic setback, life not only recovered but rose to new heights, always advancing in both diversity and complexity. We will explore this trajectory in more detail in Chapters Seven and Eight.

Calling life resilient is an understatement. Once it gets a hold on a planet it is practically impossible to destroy. If nuclear war were to suddenly break out, it would be unlikely to kill all animals or even all humans, much less all life. Even if every citizen in every nation united in a joint kamikaze effort to wipe out all biology by nuking the planet, we would not succeed. Evolution would just get a do-over. And if the sun stopped shining, microorganisms living below the surface of the earth would survive on the thermal and chemical energy being generated by its radioactive interior. Some scientists even believe that life would persist if you blew the planet to bits; bacteria would become spacefaring passengers on tiny rocks that could eventually seed other planets.

Of course, at some point in the very distant future, all the stars and planetary processes will have exhausted their means of energy production. But if the path of progress that has been so steady and reliable on Earth is typical of life-harboring planets (which the coming chapters will argue is almost certainly the case), then by that point we should expect intelligent life to have learned to utilize the many alternative sources of free energy offered up by nature. This is an exciting line of thought, but we are getting ahead of ourselves. The point is that life is robust, so much so that it almost seems daft to regard it as accidental or anomalous rather than inevitable and fundamental.

What explains life's stunning ability to stay far from equilibrium in a constantly changing, constantly turbulent environment? How does agency allow life to control its own fate and evade death and disorder for eons? What is it that makes intelligent life so unstoppable, and for at least one

planet, so utterly transformational? Although it might appear that way to the naïve observer, the secret sauce of life is not some indescribable incorporeal force; the ingredients of agency are *information and computation*.

# 5

# Biological Information and Computation

Thinking about life from the perspective of thermodynamics has led us to a new genesis story that is all natural but nonetheless magical. We call this natural magic *emergence*. The first biological system, a single cell, began as a dissipative structure that emerged from the far-from-equilibrium conditions of a geochemically active Earth. As already discussed, this self-organized structure, like all dissipative structures, had a specific function—to facilitate energy flow and entropy production.

Abiogenesis was not the chance product of one highly anomalous molecular collision but the statistically inevitable result of autocatalytic chemical systems evolving into increasingly complex forms through dissipative adaptation and phase transitions, until eventually some critical complexity threshold was surpassed and a self-replicating unit emerged. According to this narrative, life emerges under an appropriate set of non-equilibrium conditions, which suggests that biology is not unique to Earth. However, life may be unique to planets with Earth-like geochemistry. This explains much of what had been previously unexplainable, but have we really given a complete account of the origin and nature of life in the cosmos? Is an organism, in its simplest form, just an elaborate chemical whirlpool and nothing more?

The answer is an unequivocal "No." Since the properties that truly make life special are not displayed by any other dissipative structure, we can be sure that we are still missing a piece of the puzzle. Agency and resilience to entropic decay make life distinct from all other natural phenomena, dissipative structures included. These special powers—differences that make a difference to the universe—can only be understood by introducing the concept of *information*. Above all else, it is the capacity to acquire, store, process, and transmit information that separates life from nonlife. By turning to information to explain agency and robustness, we may illuminate not only how life emerged but also where it's headed.

This approach will aid us in our pursuit to discover how significant life is in the grand scheme of things. If the objective significance of a physical phenomenon can be measured by the magnitude of its cosmic impact, then how does life measure up? Is it essentially inconsequential, or is it likely to play a considerable causal role in the evolution, development, and ultimate fate of the universe? To begin to answer this enigmatic question, we must start thinking about life as an information-processing system that calculates solutions to survival challenges—in other words, a *computational system*. This description applies both at the level of the individual and at the collective level. The biosphere of which we are a part is actually an integrated global network of information-processing, error-correcting, problem-solving inference machines that collectively keep life alive and far from thermodynamic equilibrium.

## What Is Information?

The shift of focus toward information might feel like an abrupt departure from the thermodynamic narrative we have been developing, but it is a logical continuation. Biological information and thermodynamics are intertwined in such a way that the two cannot be separated. Information enables life to continually evade equilibrium through extracting energy, replicating, adapting, and learning. So, in a sense, information itself sustains the energy flow channels that facilitate the production of thermal entropy from the planet's free energy supply. Additionally, information has a direct mathematical relationship to entropy that nicely ties together many distinct

but related concepts into a cohesive cosmic story. Clearly, we have much to explain, but let's begin by unpacking what is meant by "information," since, like entropy, it is an abstract concept that means different things to different people. We are all intuitively familiar with the concept of information because in this day and age we are completely surrounded by it. Through the internet, television, music, and magazines, we are constantly bombarded with symbols and stimuli that we cannot help but process. Visual and audio information enters through sensory channels, like eyes and ears, which relay it along nerve tracts to its final destination—the information-processing unit that we call the brain.

But information has been at work in biology long before brains arrived on the scene through evolution. In school, we are taught that every cell contains genetic information, although one need not know what DNA is to deduce this much. For example, from the observation that babies reliably grow into adults and seeds into plants, it is obvious that the development of an organism is guided by some set of embedded instructions that is there from the get-go. Additionally, the structural and functional similarity between offspring and parent implies that reproduction is a process that transmits this information. And, finally, that children always look just a little bit different from their parents and their siblings tells us that genetic information must come in discrete units, what we might call "bits," some of which must get altered (through mutation) during the self-copying process. Schrödinger recognized this and wrote about it in some detail almost a decade before the DNA molecule was discovered.

It is an interesting historical fact that the scientists who discovered DNA's double-helix chemical structure, Francis Crick and James Watson, later revealed it was Schrödinger's book *What Is Life?* that gave them the idea of what to look for. How was Schrödinger, a physicist, able to understand the true nature of biology before biologists could? His insights were a product of thinking about life from the perspective of thermodynamics and information, and it is this new perspective on biology that we will turn to now and throughout the rest of the book.

Explaining life in informational and computational terms seems reasonable enough, but to be science and not just philosophy, we must be able to quantify biological information and describe it in the objective

language of mathematics. Fortunately, the growing academic field known as *information theory* is providing the tools to do just that.

## Information According to Shannon

Information was formalized into a well-defined mathematical concept in 1948 when Claude Shannon, an engineer working at Bell Labs, published a landmark paper titled "A Mathematical Theory of Communication." This article would provide the framework that information theory would be built upon. Since Bell Labs, created by AT&T and the bringer of telephones to the world, was primarily interested in communication, Shannon wanted to understand how messages could be sent from one place to another over some communication channel with optimal accuracy. Since all channels naturally generate some degree of "noise" that partially corrupts the message encoded in the signal (remember that nature is intrinsically noisy), the receiver of the message can expect to be confronted with some amount of *uncertainty* regarding its true content.

However, Shannon discovered that it was possible to reduce the amount of uncertainty by using a few "error-correction" tricks, like sending duplicate copies of a single message. Since noise is generally caused by random perturbations, it affects each copy differently, so when the redundant copies of the message are received and combined, much of what was corrupted is recovered, and uncertainty about the original content is reduced. According to Shannon's theory, information represents a "reduction in uncertainty," and the amount of information in a signal corresponds to the amount of uncertainty that has been reduced upon receipt of the message. To put it another way, uncertainty is ignorance and information is knowledge.

To clarify this concept with a classic example, let's think about the information that is gained when one flips a coin and observes the outcome. Before the coin toss, one does not know whether it will land on heads or tails; the odds are fifty-fifty. When it lands and you look at the result, you collapse two probable states into a single well-defined one, and in doing so you acquire exactly one bit of information.

Now instead of a coin, which has only two possible states, imagine throwing a six-sided die while your eyes are closed. In this example, your uncertainty is greater because the die could potentially be in more states. When you open your eyes and observe the way it actually landed, you are reducing more uncertainty, and therefore you gain more information.

Specifically, by observing the outcome of the die roll, you acquire roughly 2.6 bits of information, but we won't worry about the mathematics. The important point is that the amount of information gained is proportional to the number of possible alternatives—or the number of possible messages that could have been sent. For this reason, receiving the first symbol of a message encoded in binary, which can be either a 1 or 0, gives you far less information than the first symbol of a message encoded in English letters, since there are twenty-six letters in the alphabet, and hence twenty-six distinct possibilities.

Although we have just started describing information, a physicist familiar with statistical thermodynamics might have already noticed that it shares some curious similarities with entropy. But do these similarities mean that the two concepts are connected? Well, the mathematical equation for Shannon's uncertainty, which measures the amount of ignorance one has concerning the information in a message, is the same expression used to calculate Boltzmann's statistical entropy, a measure of configurational disorder.

This formal equivalence, first pointed out to Shannon by Hungarian American polymath John von Neumann, is no mere coincidence. Recall from Chapter Two that the more entropic a system is—the more random, mixed up, and disordered its components are—the less certainty one has about the precise state of the system (its microstate) by simply observing or measuring its collective properties (its macrostate). The reason for this is that, relative to an organized system, there are many more ways to arrange the components of a disordered system without a corresponding change to any global feature, such as its overall appearance, temperature, or pressure.

The logical similarity between Boltzmann's statistical disorder and the uncertainty quantified by Shannon led von Neumann to suggest that his measure of uncertainty should also be called entropy, as if the term weren't confusing enough. But this was partially by design; von Neumann told Shannon that since "no one knows what entropy is, in a debate you will

always have the advantage."[1] So now we have a third type of entropy, known as *Shannon entropy* or *information entropy*, and it is a measure of the ignorance, or uncertainty, one has before decoding an encrypted message.

The connection between ignorance and entropy leads us to a great insight. If uncertainty is mathematically equivalent to disorder, and information is a reduction in uncertainty, then information acquisition must be associated with a *reduction of disorder*. This makes intuitive sense, since knowledge does not float around freely. Meaningful information must be encoded or instantiated in some physical structure, like an ordered medium or an organized system. In the next part of the book, we will see why this connection between information and entropy is really the key to understanding life's origin and evolution toward higher complexity.

In trying to quantify information, we are inevitably confronted with the problem that what we consider to be or not be information seems in some sense subjective, or at least relative. A message is only informative if it tells the recipients something they didn't already know. It would seem straightforward to say that a survival guide contains more information than a book of equal size with just one word printed over and over. If a message is full of redundancy, then it does not contain much information. Similarly, if a book has words that are chosen entirely at random, then the content would just be meaningless "word salad"—not much information there either. Definitely not as much as the survival guide, right?

But what if the survival guide is in English and it is picked up by someone who only understands Japanese? For that person, the book contains no more information than if it were gibberish. What if the survival guide contained a lot of technical jargon because it was written as a manual for trained Boy Scouts? In that case, it would hold more information if the reader were a Boy Scout instead of city dweller with no survival training or education. But what if the Boy Scout was actually the book's author? Then it wouldn't be informative at all because it wouldn't be telling him anything he didn't already know before. As you can see, when it comes to classifying and measuring information, things aren't always so cut and dried.

However, Shannon, being an engineer, had a practical solution to the problem that would greatly simplify things. To be able to quantify

information in a precise and objective manner, he decided to just ignore the *meaning* of the transmitted message entirely. Stated another way, his measure disregarded semantic (meaningful) information and focused exclusively on syntactic (symbolic) information. So, for instance, if an encoded message was received and decoded with no errors, it did not matter whether that message was a random string of letters or a novel by Shakespeare. As long as the same symbol system was used, and the number and diversity of symbols in the string of characters were the same, so, too, would be the amount of information embedded in the signal.

Although Shannon was not thinking about using information as a way to explain biology, it didn't take long for scientists with interdisciplinary backgrounds to understand its relevance to life. Biological structures and processes could be more precisely articulated and modeled using an information-based approach. Today, it is hard to think about biology without thinking about information.

## Living Systems Are Informational Systems

John Maynard Smith, one of the great British evolutionary biologists of the twentieth century, described the influence that information theory has had on biology in a 1997 interview conducted by another famous British evolutionary biologist, Richard Dawkins:

> Biology, during this century, has become increasingly concerned precisely with information. Genetics is a question of how the information is transmitted between generations. Molecular biology is concerned with how it's translated from DNA into protein. Developmental biology is concerned with how it's translated from a set of proteins into a morphological structure. And evolutionary theory is essentially about how the information got there in the first place. I think you can almost characterize the twentieth century in biology as the century in which it became the science of information in living systems.[2]

But biological information is categorically different from the type of information quantified by Shannon. In particular, it is not just syntactic but also inherently semantic. As both Maynard Smith and Dawkins pointed out some time ago, biological information is "meaningful" in the sense that it builds an organism, a robust system of structural and functional complexity, and instructs it on how to survive in its environment. From this perspective, life is biological hardware running survival software. You could say that DNA contains both a developmental and a behavioral *program*, and programs require codes that specify and implement algorithms. When a fly or a worm or some other simple creature is born, it automatically matures along a predetermined developmental trajectory and it does not have to be taught how to behave in order to perpetuate its own existence—it instinctually knows what to do. With our new thermodynamic definition of life, we could say that meaningful biological information is that which keeps an organized system out of equilibrium, away from disorder and decay.

Like all meaningful information, biological information only has value in the right context. Without the cellular machinery and biomolecules that read and implement the genetic code, the information is useless and would appear to be a random string of symbols to a clueless observer. And if an organism is placed in a physical environment that is totally unlike the one it has evolved in, its behavioral program will be as useless as random bits. To really understand the amount of meaningful content that is contained in an organism's genes, we need to know how those genes relate to an agent's function and success in the appropriate natural environment.

But with all these complicating factors, how can we possibly quantify biological information so that we can answer questions like: "Does meaningful genetic information accumulate as a result of evolution?" and "Does this type of information increase as systems increase in complexity?" Such a measure would help us determine whether life on a planet must inevitably grow more intelligent as natural selection culls genetic variation. Without understanding life in informational terms, one is unable to accurately assess its cosmic significance—how much influence it is likely to have on the future evolution of the universe.

We could certainly measure the amount of syntactic information in a strand of DNA using Shannon's methods. Since DNA represents a code that

is made from molecules known as nucleotides, of which there are only four kinds (represented by the letters A, C, T, and G), crunching the numbers tells us that we acquire two bits of information per nucleotide that we decode. If we know the number of nucleotides in a species' genome, which we can determine with genetic testing, then we can simply multiply this number by two to get a decent approximation of the number of bits of syntactic information in an organism's DNA.

While this may help answer some questions, like how much computer hard drive space we would need to store the genetic information of a certain number of individuals, it does not at all seem to capture the amount of information that we've distinguished as meaningful—information that makes a difference to an organism's survival. According to this measure, a random and useless collection of three billion nucleotides with no coding function at all holds an equal amount of information as the human genome, which codes for highly intricate biological structures, functions, and behaviors. What Shannon's method fails to capture by ignoring meaning is *knowledge*—that is, information with causal power.

Fortunately, complexity science has come to the rescue. Using the mathematical tools provided by non-equilibrium statistical mechanics and information theory, researchers are developing information measures that can capture meaning and biological relevance. For example, scientists at the Santa Fe Institute, such as Artemy Kolchinsky and distinguished physicist David Wolpert, are working on semantic information measures based on an approach known as *teleosemantics*, which assumes that the system being analyzed pursues some objective. For living systems, that objective is based on the goal of staying far from equilibrium, otherwise known as survival. Wolpert and Kolchinsky formally define semantic information as "the syntactic information between a physical system and its environment that is causally necessary for maintaining the system's existence."[3]

According to Wolpert, this information measure can detect agency because it can distinguish between living systems and inanimate ones, like rocks, as well as inorganic dissipative systems, like hurricanes and whirlpools. Semantic information of this kind is sometimes called *intentional information*, but it may be easier to think about it as *adaptive information*, or *knowledge*, since it reduces environmental uncertainty, and

therefore the ignorance, of the complex system. The Italian theoretical physicist and popular author Carlo Rovelli has taken up the task of further refining Wolpert and Kolchinsky's measure, and many physicists and computer scientists are sure to follow suit once it becomes clear that explaining agency and intention is at least partially, if not entirely, a mathematical problem.

Additional information measures applicable to biological or cognitive systems, such as integrated information and effective information, have been developed or refined by scientists at the University of Wisconsin–Madison, including Giulio Tononi, Larissa Albantakis, and Erik Hoel (now at Tufts University), as well as physicists at Arizona State University, such as Sara Walker and Paul Davies. Both of these theoretical tools quantify how much a particular state of a system affects its future state. Not only can these information and entropy measures help detect agency, but they also help to quantify a system's capacity for intelligence, and perhaps even its degree of consciousness. For this reason, we will return to the concept of *integrated information* in Part Two and Part Three.

Although these measures are only in their nascent stages, they show us that quantifying biological knowledge—information that aids far-from-equilibrium survival—can someday be a tractable task. Could they also be used to determine whether the information and complexity associated with life steadily increase during evolution? Could they be applied not just to organisms but also to other complex adaptive systems like entire ecosystems or the biosphere as a whole? What about the internet? In theory, this is quite possible, and research in this area is quickly growing in popularity, although it may be some time before such applications become practical.

But living organisms do more than just store and transmit information. Biological agents perpetually process information, or perform *computation*, meaning that the origin of life is actually the emergence of nature's first computer. Living agents may be drastically different from your laptop and smartphone in material, design, and behavior, but they, too, are computing machines.

# Living Systems Are Computational Systems

For any living organized system to maintain its far-from-equilibrium state, it must be able to solve a series of computational problems. First and foremost, it has to find ways to continually extract energy from a noisy, fluctuating, and largely unpredictable environment. Also, it must achieve this goal while avoiding threats, like predators or toxic substances. Should it fail to do either of these things, it will not live long enough to make copies of itself, and the genetic information it carries will be lost. It would not be inaccurate to characterize an organism as an embodied energy-extraction and threat-avoidance program. If we closely analyze what it takes to accomplish these critical tasks, we will start to get a sense of how computation is involved, so let's look a little closer.

To evade the three Ds—death, decay, and disorder—an agent must constantly gather and analyze information about its surroundings, and based on that information, determine an appropriate response. As environmental data pours in through sense organs, sensory input is continually converted into a behavioral output, such as an action, or in some cases, inaction. This happens in much the same way that your laptop produces a preprogrammed response, like launching an application or displaying text or pictures, whenever you type in the right keyboard command or use the touch pad. For the input to trigger the correct output behavior, the living system must contain within it a stored program that implements the proper algorithms. Biological computation is called *cognition*, and cognitive ability is generally what we are referring to when we talk about intelligence. Cognition is typically associated with brains, but genomes also encode behavioral programs that generate goal-directed activity.

Let's further clarify the computational nature of cognition with some concrete examples. As complex systems that must stay far from equilibrium, humans must constantly overcome the challenges associated with finding food, our source of chemical energy. It all happens so seamlessly that it is easy to forget there are complicated computational processes occurring at every moment. Driving to the grocery store isn't a very exciting example, so let's imagine primitive man chasing a rabbit through the forest. His brain receives sensory input, the stream of visual

images of the rabbit and forest and the sounds of movement, and it computes the proper response action. He does all this without much hesitation or conscious reflection because he instinctually knows how to react. By following an algorithm—a set of decision rules that the cognitive system has inherited or acquired through experience, in this case to help meet its energy demands—he is able to roughly anticipate or predict the actions of the agent that serves as food.

The computational work in this example is carried out by the brain's complex network of neurons. But energy-seeking behavior is present in all forms of life, even those that lack a central nervous system. We've already mentioned that many plants have a tendency to move toward light, known as heliotropism, and some species even release toxins when they sense a harmful insect or parasite. These output behaviors essentially work like a reflex, and a reflex reflects an algorithm, an algorithm that is implemented when triggered by an input.

An even simpler organism that computes is the bacterium *Vibrio cholera*, which swims toward chemical food and away from toxic substances via chemotaxis. The single-celled agent has a body and a little tail called a flagellum that acts as a motor, and that's about it. When it senses chemical fuel around it, it slows its motor and feeds. When it detects no food, the tail wiggles, and the motor moves the organism in random directions until food is sensed once again. In these examples, the brainless organisms are themselves cognitive systems that compute solutions to the problem of staying far from equilibrium, using the built-in logic of life. The ability to make different decisions based on different conditions is a very primitive form of *control* and a basic requirement for agency.

A system that can exercise control is known as a *cybernetic system*. The field of cybernetics is the study of these systems, often described as teleological for their goal-directed, purposeful behavior. The term *teleology*, or purpose in nature, goes all the way back to Aristotle. Teleology was once thought to be driven by a supernatural or conscious force guiding a system toward some higher goal, but the field of cybernetics showed that goal-driven movement could be explained in mechanistic terms. A goal represents an attractor in a "fitness landscape" that a system tends to evolve toward, and teleological movement is the result of the information-processing capabilities of biological systems. However, because of its early

mystical associations, today's scientists are more comfortable talking about agency than teleology, though the concepts are essentially one and the same.

A clear definition for cybernetics can be found in the title of MIT mathematician Norbert Wiener's 1948 book *Cybernetics: Or Control and Communication in the Animal and the Machine*. Emphasis was placed on regulatory mechanisms like positive and negative feedback. With feedback you get causal loops, such that the outcomes of actions become input for subsequent actions. By utilizing feedback loops, cybernetic systems can quickly amplify processes such as those underlying self-control and self-renewal. Around the same time that Shannon was formulating his theory of information, Wiener and his European colleagues, including Ludwig von Bertalanffy and W. Ross Ashby, were developing the scientific concepts of cybernetics. We will return to the topic of cybernetics in Part Two.

While a bacterium's control process seems simple, it is only possible because the organism has embedded within its structure some *internal model* of its surroundings, albeit an extremely rough sketch in this case. This internal model can be thought of as an abstract map of the world that contains representations of the relevant environmental regularities. For agents with brains, we call this internal model a *mental model*. But even organisms like bacteria that lack the intricate memory systems of the brain are born with genetic memory that allows them to "know" just what to avoid and what to eat or absorb. Generally speaking, an organism's model of the world is what enables it to stay far from equilibrium, and in the next chapter we will learn how this model comes into existence.

In this chapter, we've demystified much of what seemed mysterious about the concept of biological agency by defining life in computational terms. Goal-directed behavior is a result of the semantic information or biological knowledge encoded in genes and brains. These answers give rise to new questions. To give a complete account of life's origin, we must explain how a dissipative chemical system evolves into an adaptive system. How did biological information, computation, and predictive ability emerge through the thermodynamic processes we've identified as the mechanisms underlying abiogenesis? And once information-storing thermodynamic machines emerged, how and why did that computational machinery evolve

into all the different species we see today, including the one with members intelligent enough to be asking how and why they exist?

In Part Two, we will not only explain why such events happened but also why they *had to happen eventually* given the conditions that were present. The emergence of cognition, and even intelligence, was just as inevitable as the emergence of life. While this statement may sound mystical on the surface, the skeptic should be assured that the underlying explanation is entirely mechanical. The destiny of life is shaped by the emergence of non-equilibrium attractors that continually create complexity by constraining chaos. While this explanation does not require introducing any new fundamental laws of physics, it does require new principles, new mathematical formulations, and a paradigm shift in how we think about the process of evolution and what we consider to be an evolutionary process.

The twentieth-century geneticist and evolutionary biologist Theodosius Dobzhansky, who helped modernize Darwin's theory by mathematizing it and uniting it with genetics, famously said, "Nothing in biology makes sense except in the light of evolution." While this is certainly true, as biological diversity and complexity can hardly be understood without the concept of natural selection, it is equally true that nothing about biology *or* evolution makes any sense except in the light of thermodynamics and information. If, as this chapter has argued, biological knowledge is life's solution to the menacing problem of entropic decay, this would suggest that evolution is a process of knowledge creation. This simple idea will lead us to a unifying theory of cosmic evolution that elegantly explains life's origin, development, and ultimate future.

# PART TWO

# EVOLUTION

*As selection carefully culls and prunes biological variation*
*Nature applies her filter to the fruits of gene mutation*
*Through the brutal and swift removal of the errors at adaptation*
*Life learns from death what not to do—and that's knowledge creation*

# 6

# Natural Selection and Knowledge Creation

---

> The digital revolution under way today is merely the latest in a long line of information-processing revolutions stretching back through the development of language, the evolution of sex, and the creation of life, to the beginning of the universe itself. Each revolution has laid the groundwork for the next, and all information-processing revolutions since the Big Bang stem from the intrinsic information-processing ability of the universe. The computational universe necessarily generates complexity. Life, sex, the brain, and human civilization did not come about by mere accident.
>
> **—Seth Lloyd, *Programming the Universe***

---

Welcome to Part Two of *The Romance of Reality*. The major lesson learned so far is that the laws of nature not only allow for life, in those places where chemical conditions are ripe, they also *necessitate* life. We still have a lot to learn, but the concepts that have been established will

allow us to understand how and why the universe is literally coming to life through a continual process of cosmic evolution.

By redefining life in thermodynamic terms—organisms are open, far-from-equilibrium, energy-cycling systems called dissipative structures—we have explained much about why biological form and function readily emerged out of early Earth's geochemistry. Although insightful, this approach hasn't really explained much about agency—the purposeful behavior that propels organisms toward survival goals.

As explained in the previous chapter, philosophers call this purposeful behavior "teleology," but for all practical purposes, teleology and agency are the same concept. If we can explain the origin of agency, we can understand how and why purpose comes into existence. That is, we can *naturalize* teleology. In the prior chapter, we learned that agency is made possible by information processing, not some shadowy supernatural spirit animating the organism—like Aristotle's "entelechy" or Henri Bergson's "élan vital." But this tells us little about how agency originated, or how it got built up into something as rich as human agency. Might the mechanisms underlying the appearance of teleology in simple life shed some light on the mystery that is free will?

## Characteristics of Agency

Since we are treading in murky philosophical waters, let's be extra clear about what we mean by "agency," and we can do that by reiterating the *behavioral* distinction that separates inanimate and animate physical systems. As we already know, inanimate things, like rocks, will always move downhill due to the force of gravity unless some external force acts on the system to oppose that motion, like a gust of wind. A high school physics student can easily compute the path of such an object using Newton's laws of motion, provided they have all the information about the different forces acting on the system. An organism, on the other hand, moves along a path that cannot be predicted using Newton's equations. Living things can climb uphill or swim upstream, *against* the force of gravity, in a quite magical manner. Life does not violate the laws of physics in any way, but it is no slave to them either. The causal power we see in the

biological world has long puzzled scientists and philosophers who can't square the intentional behavior of agents with the deterministic "billiard ball universe" envisioned by Pierre-Simon Laplace, the eighteenth-century French mathematician, astronomer, and physicist. Laplace argued that if everything in the cosmos (including life) is made of atoms that follow Newton's laws of motion, then organisms are no freer to choose their future trajectory than billiard balls are to choose their path along a pool table. This book aims to prove Laplace wrong.

While it is clear we have not yet fully explained life's origin and nature, we are undoubtedly moving closer to that goal. Now our question is not only more targeted and specific but also expressed in a language that is more amenable to the quantitative methods of modern science. "How did life emerge?" has become "How did biological information and computation emerge?" In other words, what process transformed a dissipative structure into an autonomous system that can act on the world and shape its environment? An adaptive agent of the simplest kind, like a bacterium, may not have any conscious awareness of the goals it instinctually pursues. But make no mistake, a computational or cognitive process is generating energy-seeking behavior that is purposeful, and in a very rudimentary sense, intelligent.

We usually associate such capabilities with brains and minds, but we know that this behavior emerged even before life went multicellular. Could the origin of agency be the key to understanding the emergence of free will *and* consciousness—two of the great scientific mysteries of our time? If we want to solve these age-old existential puzzles, we must start thinking about them from the computational perspective of the modern age. When we ask how inanimate matter becomes animate, we now understand that what we are really asking is how a chemical system begins processing, storing, and integrating information. Despite the progress it represents, this question has scientists so stumped that astute creationists often cite the problem as clear evidence for supernatural intervention.

But rest assured, we need not look beyond science to explain the mystery. The concepts of complexity science that have been introduced—derived from non-equilibrium thermodynamics, statistical mechanics, and information theory—arm us with all the explanatory tools we will need to succeed in our mission. And in explaining the mysterious origins of

information and computation, we will also begin to address a larger question that really is the mother of all existential questions: How significant is life in the grand scheme of things? Is it likely to influence the large-scale development of the cosmos? According to Boltzmann's statistical interpretation of the second law of thermodynamics, all forms of organized complexity—life being no exception—are destined to decay as the universe gradually moves closer to equilibrium. Surely, any human with a pulse would want to know: Is universal doom and gloom unavoidable, or can life intervene and carve out a more hopeful cosmic trajectory?

## The Need for Knowledge

To have the causal power required for even attempting such a colossal feat, a sentient system—a system capable of sensing and perceiving the world—must possess sufficient *knowledge* about how the universe works.

It is only through knowledge, and in particular knowledge of nature and natural law, that life may overcome the vast mountain of challenges associated with achieving cosmic domination. We generally acquire this kind of knowledge through a systematic procedure that entails formulating and testing hypotheses, and we call this learning method, you guessed it, *science*. No one knows for sure how far science can carry life in the cosmic battle against disorder, but some of our most respected scientific minds, both past and present, believe that its potential could be practically without limit.

For example, the 1963 Nobel Prize–winning physicist Eugene Wigner published a classic paper titled "The Unreasonable Effectiveness of Mathematics in the Natural Sciences," which reflects on the seemingly miraculous, if not godlike, power that physics has bestowed upon man. In the famous essay, Wigner remarks that "the enormous usefulness of mathematics in the natural sciences is something bordering on the mysterious and there is no rational explanation for it." Sixty years later, another visionary quantum physicist, David Deutsch at the University of Oxford, similarly spoke of the infinite reach of scientific knowledge, and the curious cosmic potential it creates for intelligent life.[1]

Deutsch is the father of quantum computing, a technology that is expected to transform the world in the coming decades. He is also considered to be the leading proponent of the many-worlds interpretation of quantum mechanics. The concept has become increasingly accepted by mainstream physics in recent years, attracting high-profile science communicators like Max Tegmark and Sean Carroll. In a short but stirring 2019 TED Talk provocatively titled "After Billions of Years of Monotony, the Universe Is Waking Up," Deutsch finishes his refreshingly optimistic speech with some inspiring words: "If one can speak of a cosmic war, it's not the one portrayed in those pessimistic stories [referring to religious doctrines]; it's a war between monotony and novelty, between stasis and creativity. And in this war, our side is not destined to lose. If we choose to apply our unique capacity to create explanatory knowledge, *we could win*." The influential American philosopher and cognitive scientist Daniel Dennett, a skeptic through and through, shared this talk on social media with a caption that said, "David Deutsch is not right about everything, but he's right about this."[2]

Similar sentiments have been frequently expressed by other eccentric geniuses, like futurist Ray Kurzweil of Google who believes that the accelerating rate of technological progress so obvious from principles like Moore's Law[*] reflects the unconstrained potential and downright inevitability of scientific progress. These sorts of reliable trends have inspired a growing number of scientists to claim that the future evolution of the universe cannot be accurately predicted without taking into account the future of intelligent life. Among the first of these mavericks was the late, great theoretical physicist Freeman Dyson, who did not mince words when, in the 1980s, he wrote, "It is conceivable that life may have a larger role to play than we have yet imagined. Life may succeed against all odds in molding the universe to its own purpose."[3]

But knowledge isn't only something that life needs in order to achieve the lofty goal of controlling the cosmos. As we have learned, the universal tendency toward disorder poses a constant threat to life from the very moment it emerges on a planet. To stay organized without violating the second law, life must be able to continually extract the free energy needed to remain far from thermodynamic equilibrium. By turning free energy into

waste heat, or thermal entropy, life exploits a loophole in Boltzmann's second law that can allow it to sustain local cosmic order for an indefinite amount of time. However, since all natural environments are constantly fluctuating and changing in a largely unpredictable manner, energy capture is no easy task. To get the free energy needed to survive, even simple organisms must possess a crude internal representation, map, or model of their environment that enables them to find energy and avoid threats. How exactly does a creature as basic as a bacterium acquire this knowledge and gain mastery over matter? The answer to this question will reveal that evolution, learning, and the scientific method all create knowledge using the same general mechanism, and they are therefore part of a single cosmic process that continuously creates complexity.

# A Unifying Theory of Reality

In this part of the book, we will answer that question, and begin to tackle the larger question of life's cosmic significance, by proposing a theory of knowledge creation that integrates thermodynamics, information theory, and evolutionary theory into a single conceptual framework. The result of this integration will be nothing less than a unifying theory of reality that explains complexity, cognition, consciousness, and cosmic evolution. Claims as hefty as these should immediately make the reader skeptical, which is the intended psychological effect. At this point in the story, we need to be thinking critically, and we need to be thinking about the importance of thinking critically. There are a lot of scientists out there claiming to have some kind of "theory of everything." What makes the idea in this book different, and not just different, but superior in the eyes of a rational skeptic?

The ideal theory of everything should not only explain the behavior of fundamental particles and forces but also all natural phenomena, including those that are emergent, like life, mind, culture, science, and technology. If the unifying theory is correct, we should see *consilience*, or converging evidence from independent fields of science. Biologist E.O. Wilson described consilience as a "unity of knowledge," and it is this unity that we

are after. Speaking of knowledge, of equal importance is that the theory be based on a sound *epistemology*.

Epistemology, the study of knowledge, is a criminally undervalued branch of philosophy. The epistemologist does not just ask what knowledge is, but also how it is acquired and how it can be determined to be valid. In the healthiest way, epistemology forces us to question our interpretation of reality. Are we logically justified in believing the things we believe? Often, we are not and never come to realize it simply because we don't bother to pose the question to ourselves.

By virtue of satisfying the concerns of epistemology, a *unifying theory of reality*—a theory of everything that includes the biological, psychological, sociological, and technological—immediately assumes a position of privilege in the space of possible scientific explanations. Why? Because it is built on the exact same principles that *define* good science. Let us explore the topic of epistemology further, as it will explain this statement and lead us right to the unifying theory we seek.

## The Problem of Epistemology

How do we acquire knowledge about reality? Initially, the question sounds straightforward, but implicit in the question is another question. What is knowledge, really? It would seem that a belief about the world, the way it is or the way it works, only counts as knowledge if it is, at least in some approximate sense, correct or accurate. By definition, beliefs proven to be false would not be knowledge. But how can we know with any certainty if a belief we hold is true or not? Don't most of us assume we are right or justified in believing the things we do? In other words, how do we know what we know is knowledge? Or is there simply no way to know?

This fuzzy philosophical problem is so fundamental that it can appear trivial, though it is anything but. Just think about what the overwhelming majority of humans considered to be the ultimate source of knowledge throughout most of recorded history. Up until very recently, just about everyone turned to religious books to explain the world around them in a collective effort to make sense of a complicated, confusing, and largely unpredictable reality. The data suggest most of us still do. The crucial

problem with organized religion is that there are as many different holy books as there are cultures, all with different answers to the same spiritual questions, different explanations for the same phenomena, different interpretations of the same events, different lessons, different teachings—you get the idea. Since logically all can't be true, or at least not in any precise sense, then some of the information in religious doctrines must not be "knowledge." That description is ideally reserved for information that fits the facts. For the longest time, humanity simply had no system for judging the validity of explanations. What makes one theory, philosophy, or worldview better than another? How do we know what to believe?

While this highly abstract but also highly practical question had already been considered in impressive detail by ancient Greeks like Plato and Aristotle over two thousand years ago, epistemology did not become a formal branch of philosophy until around the nineteenth century, emerging as a product of the Age of Enlightenment. It was the culmination of two consecutive centuries' worth of groundbreaking thought by the great pioneers of modern philosophy and science, like René Descartes, John Locke, and Immanuel Kant in the seventeenth and eighteenth centuries. As rationalists and empiricists, they realized that the only information we could really trust was that which we receive directly through our own senses. They saw valid knowledge as coming from observing reality by way of personal *experience*, a mode of understanding that is inherently qualitative, subjective, and firsthand.

This was a significant advance in our understanding of understanding, but even Descartes knew there was a serious problem with sense data. History has shown time and again that our senses can deceive us, and often in no small way. The earth is round, but from our perspective, it sure doesn't *look* like it. This ball we are all on is flying through curved space as it orbits around the sun, but it sure doesn't *feel* like it. That the sun seems to rise and fall around a still Earth is proof that appearances and experiences can't always be trusted. In fact, trusting deceptive appearances creates the inverse of knowledge, which is ignorance. Believing that the earth is flat is a timely example of such ignorance. This point nicely brings us back to Shannon's information theory, a paradigm that can provide some much-needed clarity to our epistemological mystery: *Knowledge is the*

*information we acquire that reduces our uncertainty or ignorance about the world*.

This approach sounds promising, but if we can't reliably trust our own senses, then how are we supposed to know with any certainty whether a belief or theory has actually reduced our uncertainty? Is there a solid way to separate the knowledge from the nonsense? Long ago, these kinds of difficult problems led some philosophers, suitably named "the skeptics," to give up on solving the problems of epistemology, and in their defeat, to declare that there is no such thing as objective knowledge, truth, or certainty. This attitude lives on in the postmodern philosophy that is trendy in many circles today, and its adherents are similarly misguided. While they make a point well worth noting, they fail to see why there is at least one type of knowledge that we know for sure approximates an accurate description of reality.

We can be certain that the knowledge we acquire through the practice of science is genuinely reducing our uncertainty about the world because it allows us to make increasingly accurate and precise predictions, confirmed by observation and experimental data. As we've established, knowledge is "information with causal power," and scientific knowledge certainly has causal power like no other form of information we know of. It has eradicated lethal diseases, built a global communication network called the internet, and created weapons of apocalyptic power.

Most impressively from a cosmic perspective, science takes life to unthinkable places, like outer space, and the moon, and perhaps ultimately anywhere we want to go. They say that "the proof is in the pudding"—well, science did in three hundred years what religion and pure philosophy couldn't do in three thousand. Absolute certainty and truth may always be just out of reach, but there is a practical method that moves us incrementally closer to that ideal. We call it *the scientific method*, and it instills us with powers that would look like great sorcery to anyone who lived before the Enlightenment period.

The first person to really see that science held the key to the problem of epistemology was Karl Popper, an Austrian-born British philosopher who could justifiably be called one of the twentieth century's most influential thinkers, even if much of that influence is opaque. Popper was one of the first to identify, or at least articulate, the general learning mechanism that

made science successful, but even he didn't anticipate what else that mechanism would explain: the evolution of life, culture, and the cosmos itself.

## Karl Popper's Epistemological Epiphanies

Before Popper, we knew that science worked, but we didn't really think to ask why or how. Clearly, he knew this was something worthy of deep thought, having dedicated most of his professional career toward it. He felt that science was so critical to understanding knowledge that he wanted to redefine epistemology as the philosophy of science. Some scientist-philosophers with an appreciation for epistemology, such as David Deutsch, believe he succeeded in that.

Living through the dawn of Shannon's theory and the development of cybernetics, Popper not only spoke the language of the traditional sciences, including physics, chemistry, and biology, but also the language of information and computation, which suited the logical and disciplined approach of modern philosophy well. It was this broad and up-to-date knowledge base that allowed him to see connections that would have simply been invisible to philosophers before his time. Since we have already learned that information and computation are closely tied to thermodynamics and entropy, we will incorporate these concepts into Popper's paradigm when appropriate.

Popper's grand solution to the puzzle of epistemology was right there in the statement of the problem, staring everyone in the face, but no one saw it because they didn't really know what they were looking for. The problem of how knowledge is created is no longer a problem when you realize that *problems create knowledge*. Let's unpack what we mean by this loopy statement. Problems, according to Popper, are a basic truth of life, and the need to solve them is what motivates science and scientists: "We are always faced with practical problems; and out of these grow sometimes theoretical problems; for we try to solve some of our problems by proposing theories."[4]

From this one sentence we can already see that science is a procedure for solving problems that works by generating possible solutions that we call hypotheses, theories, models, or conjectures. Because no one is psychic or omniscient, some of these theories will turn out to be right and others wrong. That is to say, some possible solutions will in fact be *errors*. But the theories that are not in error, which provide useful solutions to our problems, contain information that has proven itself to be not just information, but actual knowledge. We can think of science as accumulating *evidence-based* information, which allows us to see statistical patterns and trends in nature that we can use to make increasingly accurate predictions. It would not be inaccurate then to say that the function or purpose of science is to generate *predictive knowledge*.

We could also characterize science as a process of *inference*, meaning scientists draw logical conclusions about the way the world works from the evidence-based information acquired in the past, but with an ever-present awareness that information can and will lead to new conclusions and deeper understanding. This is the best we can do in the face of uncertainty. The fact that science could dependably produce reliable knowledge—in the form of theories or models that accurately predict the lawful or probabilistic behavior of the world—told Popper that science uses some kind of formal method to solve problems, unlike guessing or relying on gut instinct. In other words, the scientific method is an *algorithm*, a specific set of rules that can be followed to find a solution to a certain type of problem. A food recipe and multiplication rules are examples of algorithms used in everyday life, though we typically think of algorithms as the sets of instructions that computers use to do all the amazing things they are capable of.

Popper called the algorithm used by science *conjecture and refutation*, but most of us know it simply as hypothesis testing. You have a problem that you believe can be solved, you make an informed guess or conjecture, then you see if your theory can be refuted—proved false—by testing its predictions (along with exposing it to criticism). In essence, when we do experiments, we are letting reality determine which theories are solutions and which are errors, so we can think of science as a form of "reality testing." Let's use the concept of configuration space introduced in Chapter Three, also known as *phase space* or *state space*, to better visualize how this problem-solving algorithm works.

On an abstract level, we can imagine a certain practical problem as a challenge, one with a solution that exists somewhere out there in the space of possibilities, just waiting to be found by someone sufficiently motivated and clever. It may take some time, but if the possibility space—the space of possible solutions—is continuously explored *in an efficient way*, eventually a solution will be discovered. This abstract depiction isn't just an abstraction. In the field of machine learning, "evolutionary algorithms" work precisely in this way, and they are utilized in some form or another in many of today's most powerful artificial intelligence (AI) systems. In these applications, the method is commonly described as "generate and test" rather than "conjecture and refutation," but the procedure is one and the same. Possible solutions (guesses) are generated until an actual solution is found, and these solutions accumulate in some memory store while the errors are filtered out and forgotten. In science, the solutions are theories or models that accurately predict some natural phenomenon, and the successful ones accumulate in peer-reviewed journals and textbooks.

Now don't go assuming these solutions are leading to some kind of utopia where science has solved all our problems. According to Popper, our solutions will *always* lead to new problems, because explaining more about the world naturally brings new questions undreamed of before. But we should not let the never-ending chain of future problems get us down. Why? Because every solution gets retained in collective memory: in brains and books and technologies and web pages. In this way, predictive knowledge steadily accumulates, our collective uncertainty about the way the world works diminishes, and science creates genuine progress. To repeat, this progress is not an illusion, nor is it subjective or relative, that is, only progress in the eyes of a particular culture. Scientific progress is objective in the sense that it extends the causal influence of agency over matter, and it does this by giving an intelligent society a general framework or model for understanding—and exploiting—the behavior of the natural world. In theory, this progress is quantifiable, as an increase in knowledge equates to a reduction of ignorance, which we know from information theory can be described mathematically as a process of [Shannon] entropy reduction.

It is important to recognize that although science is a rule-based procedure, it is very much a creative process. A conjecture is a philosophical invention, cooked up rather mystically by the mind through

the mental computation we call careful contemplation. However, until the hypothesis is tested against reality, it is not yet truly knowledge; it is just information that represents speculation. Knowledge is information that has demonstrated its usefulness. It is what is left over after cycles of experimental testing have eliminated false theories. As scientists continually test their hypotheses and tweak their models to account for new and surprising data, a kind of "learning loop" emerges that statisticians call *Bayesian updating*. Based on Bayes' theorem, developed by eighteenth-century English statistician, philosopher, and Presbyterian minister Thomas Bayes, Bayesian updating refers to a mathematical process whereby an accepted theory or predictive model gets increasingly accurate through the iterative testing of competing variants of that theory.

Updating one's model corresponds to updating one's certainty regarding previously held beliefs, and Bayes' theorem gives us a formula for calculating exactly how our confidence should change based on new findings. While the growth of scientific knowledge as a whole follows this equation only in an informal, sloppy, approximate sense, the metaphor makes it easy to see why technological progress occurs so reliably. Science is far from infallible, but because each field of study is constantly self-correcting and updating its "beliefs" about the world, the growth of knowledge steadily shrinks the sea of ignorance. According to Popper, knowledge does not come from observation, but from experimentation, and a practical or theoretical problem always provides the motivation.

This appears to be a satisfying answer to the problem of epistemology—we gain knowledge by formulating and testing theories that solve our practical and intellectual problems—but it instantly leads to a new conundrum. Popper's solution, conjecture and refutation, seems to imply that knowledge can only be acquired by self-aware beings intelligent enough to generate logical theories about how the world works. Now, that seems awfully anthropocentric, not to mention at odds with the narrative that knowledge acquisition is a universal biological imperative. Obviously, humans knew quite a bit about how to live in nature and society successfully long before the scientific method or the logic of inference existed. What explains all the practical knowledge that humans, or for that matter all animals, use to solve their basic survival challenges?

Let us not be worried by this conundrum, as Popper's epistemology predicts that solutions will always lead to new problems, and that applies to his own solution to the problem of epistemology! Progress requires the continual emergence of new problems, and this particular philosophical problem led Popper to his next epiphany: the scientific method is a refinement of the same general method we use to solve problems all throughout life.

## Adaptive Learning = Hypothesis Testing

Once Popper understood that the knowledge-generating mechanism behind science's success was conjecture and refutation, he realized that human learning, which begins at birth and continues till death, uses the same problem-solving algorithm, although the developmental psychology literature called the method *trial and error*. Because life is constantly presenting us with new challenges—such as the need to get from one place to another, a problem that inspires babies to learn to walk—we must constantly try out new behavioral solutions. We can think of these actions as "guesses about how to survive," or if you prefer, "experiments in evading equilibrium," or even "predictions for persistence"—and they will often fail.

Our instincts may nudge us in the general direction of a solution, but an instinct alone is not going to get us exactly where we want to go, at least not on the first try. When our initial behavioral conjecture does not get us to our goal, we try again, this time with a new "theory" that is essentially the old one but with a little twist, tweak, or tinker. If it works, the behavioral solution is remembered and stored for future use, ready to be repeated but also revised when novel situations present themselves, as they often do. In this way, we learn from our errors. It is a bit like that line from Christopher Nolan's *Batman Begins*, when Alfred tries to inspire Bruce Wayne not to give up after failing: "Why do we fall, sir? So that we can learn to pick ourselves up."

Like hypothesis testing, the trial-and-error method of problem-solving explores the space of possible solutions in a manner that is in some sense "blind"—we do not know the correct solution to a problem in advance—but

still systematic and therefore efficient, as we naturally begin with those potential solutions that are closer to our starting point in the solution space we are sampling. For example, if a baby reaches for its bottle and barely misses, it will adjust its behavior only slightly, which minimizes the chance of making an error.

Even when we get the idea for a new behavior from watching another person, as in "imitation learning," there is still some degree of trial and error involved, since executing a new motor command is a miniature experiment in itself. Not to mention the behaviors we learn from others were all originally created through trial and error at some point in history. We call the knowledge acquired through this implementation of the universal learning algorithm *adaptive behavior*, and like scientific knowledge, it accumulates through the elimination of errors and the selective retention of solutions to practical problems.

As the brain encodes adaptive information in the synaptic circuitry created by networks of neurons, our mental model of the world, the representation of reality that defines the mind, gets built up and updated continuously through a process roughly analogous to Bayesian updating. One of the trendiest theories in neuroscience today is called the *Bayesian brain hypothesis*, championed by the world's most influential neuroscientist, Karl Friston of University College London. According to the alleged "unified brain theory," sentient systems naturally try to resolve environmental uncertainty by behaving in ways that *minimize prediction error*—or the amount of "surprise" an agent experiences in its encounters with the world—which makes the mental model more accurate. In less technical terms, without even thinking about it, we constantly make predictions about what is going to happen next, and we self-correct when those predictions don't pan out, making us better predictors over time (ideally). For this reason, just like science, adaptive learning is a knowledge-producing process that updates and expands life's model of external reality.

Although the Bayesian brain hypothesis is quite abstract, we should know it is not just some purely speculative, untestable, pie-in-the-sky theory. In addition to the growing body of supporting evidence from neurophysiology and experimental psychology, a Bayesian approach to machine learning is producing AI systems that can outperform our most

powerful “deep learning” algorithms when it comes to a wide variety of practical tasks. Though the mathematical formalization and digital implementation of the idea is new, the characterization of the brain as an organ that performs statistical inference (i.e., “prediction”) can be traced back to the work of Hermann von Helmholtz, a trailblazing German physicist and physician who saw the connection between adaptive learning and hypothesis testing almost a century before Popper articulated it more formally.

If you are finding it hard to believe that invisibly embedded in that squishy thing inside your skull is a detailed model of the outside world, simply close your eyes and try to imagine the room you are in. It’s there immediately, and impressively accurate, wouldn’t you say? Now try to zoom out and see the neighborhood you live in from a bird’s-eye view. Next, imagine how Earth looks from outer space. This simple task reveals just a speck of the complexity of the world model that has been built up by your brain over a lifetime of learning. Your ability to survive in a complex and challenging world, indeed your very existence, is proof that you have at least *some* knowledge, even if you also possess many unjustified beliefs, which are all those that have not yet been tested.

Of course, you don’t *have* to test your beliefs to get through life, but I wouldn’t recommend this strategy. They say that “ignorance is bliss,” but not when the real world is hitting you with practical problems left and right. If you don’t want to hide in a dark room your whole life, avoiding any new information that might challenge or disconfirm your chosen worldview, then you better be willing to either update your beliefs in the face of new evidence or live with constant cognitive dissonance. Popper was clever for recognizing that both science and adaptive learning generate knowledge through the same mechanism: iterative hypothesis testing or trial-and-error elimination. But, once again, this explanation leads to a new knowledge paradox. How was life able to acquire knowledge before brains emerged? Doesn’t learning require one of those things? We typically think of biological memory as a unique feature of neural machinery, making it difficult for us to conceive of a learning mechanism existing prior to central nervous systems, but even supremely simple life-forms like bacteria and plants show goal-oriented behavior, which suggests some sort of basic survival program encoded into memory. The problem of epistemology

brought Popper to the problem of teleology, which is identical to the problem of agency we've been trying to answer. But this problem led him to his greatest epiphany yet, and it is taking us there too.

## Evolution Is a Knowledge Creation Process

The answer to this paradox is of course DNA, life's original memory system. Unlike memory stored in our brains, which acquires new information through lived experience, genetic memory accumulates in an evolving population of organisms over many generations, as a result of replication with mutation and natural selection, or *variation and selection*. Cyberneticists like Norbert Wiener were characterizing adaptation as "phylogenetic learning" as far back as the 1940s, and this undoubtedly had a strong influence on Popper's thinking. In biology, *phylogeny* refers to the evolutionary history of a species while its opposite, or complement rather, *ontogeny*, refers to the embryological or cognitive development of the individual.

Ontogenetic learning, the kind of learning we are most familiar with, occurs during an organism's lifetime; in fact, you are doing it right now as you learn about ontogenetic learning! We've already learned that ontogenetic (i.e., adaptive) learning occurs through the mechanism of trial and error, which is equivalent to the algorithm of scientific hypothesis testing: conjecture and refutation.

Phylogenetic learning, on the other hand, occurs through variation and selection and takes place at the level of the species over generations. Popper's greatest insight would come from recognizing that all these mechanisms were actually one and the same, although this realization wouldn't come all at once.

First, Popper saw that the growth of scientific knowledge mirrored an evolutionary process, and in a way, that seemed deeper than your average metaphor: "In science . . . theories are highly competitive. We discuss them critically; we test them and we *eliminate* those theories which we judge to be less good in solving the problems which we wish to solve; so only the

best theories, those which are most fit, survive in the struggle. This is the way science grows."[5]

Once Popper understood that *science was actually a Darwinian process*, it didn't take him long to see that biological evolution could also be interpreted as hypothesis testing. Ever since life took hold of the planet, the biosphere has been practicing a primitive form of science, accumulating evidence-based knowledge by testing the "embodied theories" we call organisms: "It is clear that this view of the progress of science is very similar to Darwin's view of natural selection by way of the elimination of the unfit: of the errors in the evolution of life, the errors in the attempts at adaptation, which is a trial and error process. Analogously, science works by trial (theory making) and by the elimination of errors."[6]

Truth be told, Popper was initially skeptical of Darwin's theory because it wasn't easily testable and therefore potentially unfalsifiable—for him a theory was not scientific unless it could be proven false. But when he understood that conjecture and refutation, trial and error, and variation and selection were functionally equivalent algorithms, it became clear that scientific progress, adaptive learning, and biological evolution were all part of a single uncertainty-reducing, information-creation process. As the visionary philosopher of science put it, "From the amoeba to Einstein, there is only one step."[7] Out of this grand insight a paradigm with unparalleled explanatory power would emerge.

## Evolutionary Epistemology: A Unifying Paradigm

With support from the work of the cyberneticists, Popper's conceptual unification of evolution, cognition, and science inspired the naturalistic approach to epistemology known as *evolutionary epistemology*, developed by the cognitive psychologist Donald Campbell in the 1960s and elaborated further by the Nobel Prize–winning zoologist Konrad Lorenz, and even Popper himself, in the following decades. According to evolutionary epistemology, cultural and technological evolution, even the evolution of chemical dissipative structures, could be explained as products of variation and selection, now understood to be a learning algorithm. "Life is a

cognitive process," wrote Lorenz.[8] At this point in the cosmic journey, we are ready to introduce a radical idea so sensible that once it is understood, it seems trivial: *All evolutionary processes are learning processes, and all forms of learning are evolutionary processes*.

This little statement may not sound paradigm shifting, but it has some pretty far-reaching implications. A functional equivalence between the mechanisms driving evolution, learning, and science suggests that *adaptation and scientific knowledge are actually the same thing*. Biological adaptation represents knowledge of the environment, and the knowledge we acquire through learning and science reflects adaptation to the environment. If you think about it, there really is no meaningful distinction between adaptive information and scientific knowledge—both allow life to predict an uncertain world, control matter, constrain chaos, and construct order from disorder. The very same property that is responsible for life leaving the planet in spaceships is what permits an organism to climb uphill. That property is called agency, and it emerges from the information stored in genetic, neural, or cultural memory. But how does all the functional information in genomes come from random mutation? Surely structures as complex as the heart and the eye could not have evolved through blind and unguided processes, a creationist would argue.

Mutation is in many ways random, but evolution is not. Natural selection sifts through the blindly generated variety of forms and finds functional matches between adaptive system and environment, organism and niche. One scientist who has done much to demystify teleology by explaining purpose as a product of natural selection is neuroanthropologist Terrence Deacon at the University of California, Berkeley, who said, "An evolutionary process does not merely transmit information into the future; it generates it dynamically."[9]

This general idea is present in some way or another in all of the major scientific paradigms that have tried to integrate biology and evolutionary theory with thermodynamics and information theory over the last hundred years. These cross-disciplinary frameworks have different names—cybernetics, general systems theory, evolutionary epistemology, complex adaptive systems, universal Darwinism, universal selection theory, and the currently fashionable "free energy principle." But they all explicitly or

implicitly understand natural selection to be a process that effectively transfers information from the environment into an adaptive system.

Knowledge creation through blind variation and natural selection elegantly solves the problem of epistemology and the problem of teleology by providing an intelligible mechanism for how organisms come to possess accurate internal representations (knowledge) of the external world, a mystery Plato desperately wanted to solve.

## Adaptations Encode Knowledge

Not yet convinced that adaptation is essentially the same as knowledge? Good! A good epistemologist should not be convinced so easily. Let's see how Popper's paradigm holds up to scrutiny. Recall the distinction between semantic and purely syntactic information made in the previous chapter. To count as knowledge, information must have meaning, and to have meaning, information must be *about* something. Biological knowledge is the information embedded in organisms *about how to survive and replicate in a fluctuating environment*. It is this "aboutness," as Deacon and Dennett call it, that qualifies genetic information corresponding to adaptation as semantic or meaningful information, and specifically uncertainty-reducing information—or knowledge. "What counts as useful information in biological evolution is determined after the fact with respect to its ability to pass through the functional error-correction mechanism of natural selection," Deacon wrote in his celebrated article "Shannon—Boltzmann—Darwin: Redefining Information."

Let's consider a couple of classic examples taken from the evolutionary epistemology literature. In a completely literal sense, a dolphin's streamlined design, which is a product of the information stored in its genome, contains a knowledge of hydrodynamics. Similarly, an eagle's wing design contains a knowledge of aerodynamics. Not only can we be certain that engineers see knowledge in these functional structures, there should be little doubt that these designs provided the original inspiration for our aircraft and submarine technology. Does anyone really think we'd have planes right now if evolution had not first produced creatures that could fly? We can only guess how long it would have taken these inventions to emerge

had nature not provided us with a dazzling array of blueprints to inspire our machine dreams.

But if theories are motivated by practical problems, and if an organism is a type of embodied theory, then exactly what problem is life "trying to solve" through adaptation? A biologist would call it "survival," but a cyberneticist would call it "persistence," and a physicist would call it "evading equilibrium." These all mean roughly the same thing, but if we are not thinking about an organism as a far-from-equilibrium system that requires continual energy to maintain its existence, then we will not understand the true role of information in biology.

Presuming the second law of thermodynamics holds everywhere in the universe, then intrinsic in all living organisms, all conceivable agents, is a basic cosmic imperative. By now it should be wired into our brains that the eternal challenge for any ordered system is resisting the natural tendency toward disorder. If you are an adaptive system, simply by existing, you have an existential problem you must overcome.

## Nature Selects the Most Stable Configurations

Eric Smith and Harold Morowitz described this existential problem in their book *The Origin and Nature of Life on Earth*: "In living systems, every level, from all atomic locations and bonds, to control, replication, and selection, is pervasively and continually subject to perturbing fluctuations." This means that to merely persist, much less replicate, a dynamical system must have a *stable configuration*. This notion of selection for stability gives us a new and enriched perspective on fitness. One popular criticism of Darwin's theory of natural selection, often summarized as "survival of the fittest," is that no one knows exactly what "fitness" is, because the standard definition is tautological, or circular, and therefore vacuous.

In her 2006 book *Godless: The Church of Liberalism*, the far-right media pundit and professional troll Ann Coulter capitalized on this ambiguity in a passage that is now well known among creationists: "Darwin's 'theory' is generally nothing but a circular statement: Through the process of natural selection, the 'fittest' survive. Who are the 'fittest'? The ones who survive! Why look—it happens every time!"

This argument is an obvious straw man, but it forces us to think a little more deeply about what it is that evolution is optimizing through variation and selection, if anything. Formally, "fitness" is a quantitative measure of reproductive success that can be precisely calculated for any organism in a population, but this still does not tell us much about what it means to be fit. Dawkins, who invented universal Darwinism with his concept of the "meme"—a paradigm that has become practically synonymous with evolutionary epistemology—believes fitness can be understood more broadly as being about stability: "Darwin's 'survival of the fittest' is really a special case of a more general law of *survival of the stable*."[10]

It seems we are getting somewhere, but what makes a particular configuration stable? We know that to avoid the collapse to equilibrium, a system must be able to continually absorb free energy from a noisy, fluctuating, ever-changing environment. To accomplish this task and maintain a stable non-equilibrium steady state, an adaptive system must acquire information that reduces its ignorance or uncertainty about the world it inhabits. The problem for life then is one of inference. To persist in a stable form, life must learn, and it must begin doing so the moment it comes into existence.

## Universal Bayesianism Updates Evolutionary Epistemology

"[A]ll biological processes can be construed as performing some form of inference, from evolution right through to conscious processing," said Karl Friston, whose *free-energy principle* makes the math underlying the Bayesian brain theory applicable to any conceivable self-sustaining system, be it terrestrial or extraterrestrial.[11] This principle basically says that to stay out of equilibrium, an adaptive system must engage in Bayesian-like learning to deal with environmental uncertainty.

The problem of eternal uncertainty might sound like an insurmountable barrier for life, but it turns out to be just what biology needs to evolve open-endedly. If you have an existential challenge, then you have a selective pressure, and if an organized system can explore a variety of configurations,

then it can search the possibility space to find solutions to its thermodynamic problem. Fortunately for life on Earth, there's always free energy in the environment that can be made accessible to the system, so extracting it is simply a matter of finding the right configuration through trial and error.

From the perspective that Dennett calls the *intentional stance*, which attempts to predict a system's behavior by interpreting it as an agent with intrinsic goals, we can attribute a thermodynamic goal to biological evolution: through adaptation, life performs inference in order to stay far from equilibrium. As recognized over half a century ago by the 1965 Nobel Prize–winning French biologist François Jacob, an organism is "a sort of machine for predicting the future—an automatic forecasting apparatus."[12] This Helmholtzian view of life suggests that when an evolving population adapts to its niche over evolutionary time, it is as if the genome of the species is "updating" its model of the environment to be increasingly accurate.

Since nature is complex, and to some extent intrinsically unpredictable thanks to hidden causes, deterministic chaos, and memory capacity limits, the organism's model is at some point guaranteed to fail when reality "surprises" it with the unexpected. An organism gets eaten by a predator, a baby fails to reach their bottle, a scientific theory fails to explain new data. What this means is that the model used to predict reality is not completely accurate. It contains a certain amount of prediction error that needs to be corrected.

So, what's the solution? Try something new. The old model worked well enough up to a point, so don't ditch it altogether. Just vary it a bit and see if it performs better. If it does not reduce surprise, eliminate that variant and try again. If prediction error is reduced, replace the old model with the new one, so that it becomes the reigning theory and the design template for new variants to be based upon. When a model has been updated in this way due to natural selection, adaptive learning, or experimental testing, we can say that knowledge has been acquired and environmental uncertainty reduced. By now, we can see a simple story emerging. By reducing Shannon entropy, or ignorance, life is able to extract the energy it needs to reduce Boltzmann entropy, or disorder. The growth of knowledge and the spread of organized

complexity, therefore, go hand in hand. The second law is the impetus for learning.

Of course, no model can ever be perfect since nature is noisy, so prediction error can at best be minimized, not resolved entirely. A well-adapted species is like a well-tested theory: provisionally accepted as true but subject to continual revision or even complete replacement. Deutsch wrote in his mind-bending 1996 book *The Fabric of Reality*, "The whole [scientific] process resembles biological evolution. A problem is like an ecological niche, and a theory is like a gene or a species which is being tested for viability in that niche."

To be clear, neither life nor the evolutionary process has any way of knowing the correct solution to a particular environmental challenge in advance, or even that a challenge exists, but a lack of foresight or conscious intent won't stop biology from incessantly "guessing." Thanks to the miracle of mutation—a random DNA copying error caused by solar radiation or a thermal fluctuation—self-replication inevitably generates genetic variation. As a result, each time an organism reproduces, an evolving population tries out a new functional design that may or may not be an innovation.

The production of new functional designs can be interpreted as the genome of a species searching the space of possible configurations for the survival solutions that we call adaptations. Through variation and selection, the biosphere ceaselessly experiments with novel structures, physiological functions, and instinctual behaviors. Every now and then, it stumbles upon a new solution to the general problem of staying far from equilibrium. Since mutations are random or blind in the sense that they could either help or hurt the organism, many of these "experiments in living" will fail.[13] In other words, organisms with inaccurate representations or models of their environment, which have dysfunctional or suboptimal energy-extraction and threat-avoidance programs, will die without reproducing. This is just a fact of nature: before brains, for an adaptive system to learn from its mistakes, many organisms have to die.

## Life Is a Signal Among Molecular Noise

The organisms that survive the filter of natural selection get to reproduce and make copies of their well-adapted functional design. Because better-adapted designs tend to survive longer and get to make more copies of themselves, adaptations will naturally spread in an evolving population. The resulting redundancy of functional designs preserves the general organism template while still allowing for innovation through mutation. This is not unlike Shannon's error-correction technique described in the prior chapter. A message sent across a channel will always be partially corrupted by noise, but since noise is random, each copy that is sent is affected differently, such that sending duplicate copies can allow the original message to be accurately reconstructed by the receiver. In evolution, the noise in the channel is not a nuisance but the source of the genetic variation that makes continual adaptation and novelty generation possible. At the same time, genetic redundancy preserves the growing statistical "signal" created by the persistent patterns of robust biological organization against a background of molecular noise, chaos, and disorder.

Comparing life to a signal in Shannon's information theory is not just a helpful metaphor, it is potentially a way to quantify evolution and the growth of diversity and complexity in the biosphere. Deacon explained: "The reduction of genotypic and phenotypic variety by virtue of differential survival and reproduction of certain organism forms (favored in a given environment) is directly analogous to the reduction of signal entropy in Shannon's analysis. Thus, in principle, evolved information can be quantified and rates of information evolution can be compared."[14]

In the previous chapter, we introduced the teleosemantic information measures of Wolpert, Kolchinsky, and Rovelli, and it is these types of theoretical tools that stand to make the program of Deacon and Dennett a reality. Similar formulations have been developed by the Canadian scientist John Campbell, author of the book *Universal Darwinism: The Path of Knowledge*, who has shown that evolutionary processes at all scales can be interpreted as Bayesian updating and described using Friston's free-energy principle equations. Friston and colleagues around the world have taken up the task themselves, yielding a slew of influential papers in recent years.

This means a rich and comprehensive mathematical formalization of evolutionary epistemology and universal Darwinism is not only possible, it

already exists, and we can alternatively call this quantitative theory of knowledge creation *universal Bayesianism*. While all three names correspond to the same formal theory, each characterization gives us a deeper understanding of the nature of adaptive complexity. Inference or prediction seems to be another unifying theme, just as important as evolution and learning, as it provides formalisms that reveal a unity among disciplines as diverse as neuroscience, machine learning, and the origins of life. Taken together, these theories offer all the tools needed to naturalize teleology: evolutionary progress becomes quantifiable and a physics of purpose becomes possible.

## Putting All the Puzzle Pieces Together

We are finally ready to explain the origins of biological information and computation, and we will do so with a rhyming equation to help with understanding as well as memorization:

*Adaptation = Statistical Correlation = Mutual Information = Model Optimization = Knowledge Creation*

These terms sound technical but the story they tell is surprisingly simple. In his award-winning essay, written in response to the Foundational Questions Institute (FQXi) contest question "How can mindless mathematical laws give rise to aims and intentions?" physicist Carlo Rovelli explained that a well-adapted organism is one that is statistically correlated with its environment: "A bacterium that swims to the left when nutrients are on the left and swims to the right when nutrients are on the right prospers; a bacterium that swims at random has less chances."[15]

This is an example of a behavioral adaptation, but structural adaptations also increase the correlation between a species and its niche. We all have an intuitive sense of this correlation even though we might not realize it. If you were shown a picture of a species you had never seen before, but the organism had gills and fins, rather than feathers and wings, you would be able to make a pretty decent guess about the kind of environment it inhabits. Conversely, if you were shown an empty environment, assuming it

looked at all familiar, you would have a decent guess as to what kind of creatures might occupy it. How are we able to make this inference? Because the evolving organism is shaped by its environment and is, in many ways, a reflection of its environment. A philosopher would say that a well-adapted species is isomorphic to its niche. Cyberneticists and Friston would say that the organism is a model of its niche.

A statistical correlation between the configuration of the organism and the state of the environment means that the two systems share information. Building on pioneering work by the epistemologist-philosopher Fred Dretske, who applied Shannon's theory to natural selection a half century ago, Wolpert and Rovelli call this shared information *mutual information* (also known as "relative information"). As a species adapts to a niche, both in form and function, the amount of mutual information between the organism and its environment grows, such that we can say the adaptive system "learns" about its niche. For a species, growing more correlated with a niche means gaining more knowledge about that niche, which equates to becoming a better *predictor* of that niche. Semantic information is the mutual information that is causally necessary for the organism's survival.

In his influential essay "The Concept of Information in Biology," evolutionary theorist John Maynard Smith explained how statistical correlation between two systems can create predictive knowledge, also citing the Dretske's work: "Dretske (1981) argues as follows. If some variable, A, is correlated with a second variable, B, then we can say that B carries information about A; for example, if the occurrence of rain (A) is correlated with a particular cloud (B), then the type of cloud tells us whether it would rain."

Updating this description slightly to fit in with our Bayesian view of life, we can say that *if some adaptive system is correlated with some environment, then the adaptive system stores a model of that environment*. As adaptive solutions to environmental challenges accumulate in genetic memory, the accuracy of the predictive model of the species gets optimized, environmental uncertainty is reduced, and knowledge is created. So:

*Adaptation = Statistical Correlation = Mutual Information = Model Optimization = Knowledge Creation*

Got it?

Now that we understand our evolutionary epistemology equation, we can see exactly how natural selection "transfers" information from the environment into the evolving biosphere, and why that information creates the intelligent, purposeful, goal-directed behavior that we associate with life and mind. Whether this intentional behavior is always accompanied by some degree of conscious experience will be addressed in Part Three, but for now, let's just say that we should not necessarily assume that agency implies mind. That being said, most would agree that we cannot truly understand mind without understanding the origins of agency. Going a step further, we can say that solving the puzzle of agency now is what is going to allow us to solve the mystery of mind later. Rovelli, the theoretical physicist known for his work in the field of quantum gravity, felt the need to address this issue in his essay:

> The definition of "meaningful" [information] considered here does not directly refer to anything mental. To have something mental you need a mind and to have a mind you need a brain, and its rich capacity of elaborating and working with information. The question addressed here is what is the physical base of the information that brains work with. The answer suggested is that it is just physical correlation between internal and external variables affecting survival either directly or, potentially, indirectly.[16]

But Wolpert, Kolchinsky, and Rovelli's model explains how functional information and teleology arise through Darwinian evolution, which comes *after* life emerges. So have we really explained the origins of agency and computation? Fortunately, we can apply the same logic to dissipative adaptation, the Darwinian-like thermodynamic mechanism from Chapter Three that describes how a system of inanimate particles can start exhibiting the collective energy-seeking behavior that we see in life.

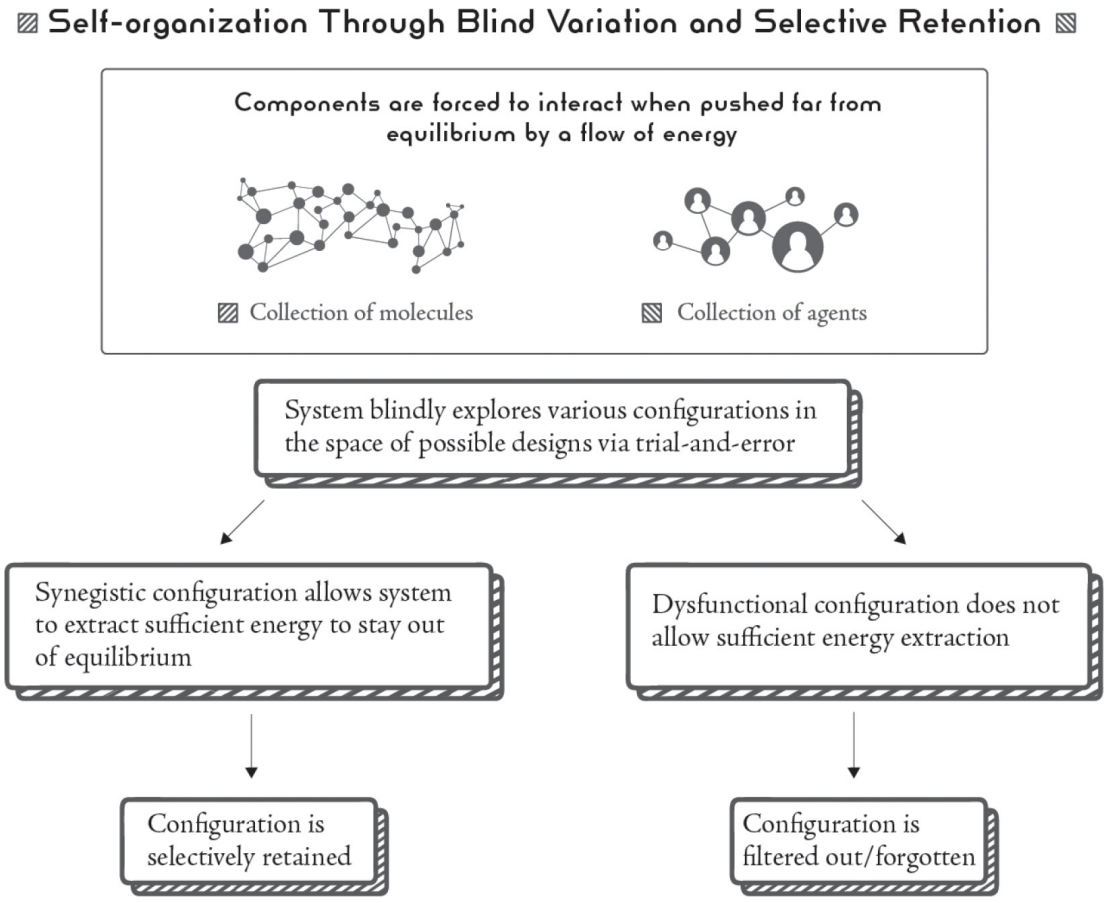
## Origins Finally Explained

Recall that dissipative adaptation is a process of spontaneous self-organization that occurs when a system of interacting molecules is driven far from thermodynamic equilibrium by the flow of energy. The principle underlying this evolutionary mechanism is as old as Prigogine, but we now have a formal model for conceptualizing how entropy maximization laws can get us from a collection of interacting molecules to self-regenerating systems (autocatalytic sets) to self-reproducing systems (organisms), and perhaps further.

Looking through the lens of evolutionary epistemology, we see that dissipative adaptation—in fact, all forms of self-organization—are ontogenetic learning processes that implement the variation-and-selection algorithm and therefore inevitably evolve toward configurations of greater stability and functionality over long enough time scales (those that persist, at least).

Pulling inspiration from the cyberneticists, social scientist Donald Campbell described self-organizing systems as adapting through an evolutionary mechanism he called "blind variation and selective retention," which is simply natural selection favoring configurations that confer stability.

In this adaptive process, the learning does not result from competition, as it does in phylogenetic learning, but cooperation. As the individual components of the system interact with each other, the elements become increasingly correlated, and the collective system they comprise explores various configurations in the state space. Through trial and error, the emergent system inevitably finds arrangements that are better at extracting energy from a fluctuating, statistically noisy stream of energy. Because states that absorb energy are stable, they are selected, and because they create entropy, they are harder to reverse. Think of dissipative adaptation as a randomly fluctuating system "getting stuck" in the most stable states serially. Jeremy England, who formalized the concept, explained that the structural organization of a well-adapted dissipative system will display something analogous to an evolutionary history of these hard-to-reverse transitions.


Self-organization Through Blind Variation and Selective Retention
Components are forced to interact when pushed far from equilibrium by a flow of energy
Collection of molecules
Collection of agents
System blindly explores various configurations in the space of possible designs via trial-and-error
Synegistic configuration allows system to extract sufficient energy to stay out of equilibrium
Dysfunctional configuration does not allow sufficient energy extraction
Configuration is selectively retained
Configuration is filtered out/forgotten

Wolpert and Rovelli's model suggests that these states should be correlated with the dynamics of the thermodynamic (energetic) environment. A system undergoing dissipative adaptation would then be, in an abstract but mathematically precise sense, mapping or modeling the surrounding energy landscape as it evolves. There is no specific place to point to and say "information begins here" because it is distributed across the collective molecular architecture of the emerging epistemological system. Knowledge comes into nature just the same way it accumulates through evolution, cognition, and science. A trial-and-error search "finds" stable solutions and records them in memory, updating the adaptive system's burgeoning world model. A 2012 influential paper called "The Thermodynamics of Prediction," authored by Susanne Still and colleagues, describes this basic mechanism before dissipative adaptation had a name. The wording of the abstract is a bit technical, but we should now be conceptually equipped to appreciate it:

> A system responding to a stochastic driving signal can be interpreted as computing, by means of its dynamics, an implicit model of the environmental variables. The system's state retains information about past environmental fluctuations, and a fraction of this information is predictive of future ones . . . Our results hold arbitrarily far from thermodynamic equilibrium and are applicable to a wide range of systems, including biomolecular machines. They highlight a profound connection between the effective use of information and efficient thermodynamic operation: any system constructed to keep memory about its environment and to operate with maximal energetic efficiency has to be predictive.

In a 2017 article, England told the digital magazine *Nautilus* that a future goal is to demonstrate "emergent computation," which he explained is all about prediction: "If we succeed in doing this, the argument will be that somehow the particles in the system are interacting in such a way as to effectively implement a calculation about the future based on the statistics of the past."

## Evolution as an Inference Engine

In the final chapter of their book, Morowitz and Smith propose a mathematical model of abiogenesis and evolution as a process of *Bayesian model selection*. But they do not restrict the process of inference to individuals—the biosphere itself is doing prediction and model updating to keep the entire living network out of equilibrium. Understanding "the nature of evolution as an inference engine" can explain things about life that were not addressed by classical evolutionary theory, like the reason for its hierarchical architecture and unbelievable resiliency and robustness.

Those working in the field of emergent computation, like complexity pioneer James Crutchfield, believe that an understanding of how nature begins doing inference could lead to dramatic breakthroughs in artificial intelligence, nanotechnology, and computing more generally. For these purely practical reasons, neuroscientists, psychologists, philosophers, and machine-learning researchers should all be interested in understanding life's

informational origins. This can be achieved with a new framework that unifies evolutionary epistemology, universal Darwinism, and universal Bayesianism under a cohesive cosmic narrative, which we will see has the power to connect life, mind, and cosmos.

So, there you have it: a Bayesian explanation for the origin of adaptive complexity (life), agency (teleology), and computation (information processing) grounded in epistemological logic and formalized with measures from information theory and statistical thermodynamics. At this point in our story, you might be thinking, This explains so much! Or, you could be thinking, There's so much this doesn't explain! Either way, you'd be right. This ride is just getting started, but we now have a solid framework for a unifying theory of reality that appears to be profoundly Darwinian and profoundly teleological at the same time.

David Krakauer, an evolutionary biologist and the president of the Santa Fe Institute—a mecca for complexity science and a home for world-class physicists and computer scientists who have taken up the task of understanding adaptive systems at all scales—articulated this emerging view of biology in a recent interview: "Life seeks to represent the world in which it lives: to encode reality. That is what a genome does. That is what a brain does. And the initial impetus towards that kind of reflection of the universe in living matter is not that well understood, I'd say."

"It sounds dangerously teleological when you say it," remarked podcast host Toby Shannan.

"Yes," said Krakauer, "It does sound that way."[17]

A Darwinian teleology certainly sounds like a paradox, but it is one we must embrace if we are to understand the evolving universe and our place in it. By turning the paradox into a principle, we can make sense of the inevitable and open-ended growth of adaptive complexity in a cosmos abiding by the second law. We can refer to this principle as "Popper's principle," which is based on a simple premise: *Problems create progress*.

---

* According to Kevin Kelly in his 2010 book *What Technology Wants*, "In brief, Moore's Law predicts that computing chips will shrink by half in size and cost every 18 to 24 months. For the past 50 years it has been astoundingly correct."

# 7

# A Unifying Theory of Reality

---

> We start with observations, and if the evolving cosmos has an observed direction, rejecting that view is clearly nonempirical. There need not necessarily be a knowable end point, but there may be an arrow.
>
> **—Harold Morowitz, *The Emergence of Everything***

---

Through our investigation into the epistemological mystery of how life acquires knowledge, we have arrived at a unifying theory of reality that challenges everything we thought we knew about the universe and our place in it. As this chapter and the next will show, this is a theory of open-ended knowledge creation and a true "theory of everything" in that it explains emergent phenomena, which can't be fully understood or predicted by the fundamental physics theories we hear so much about, like string theory, a supposed "theory of everything"—everything except the things we care about most, that is. The old idea of a comprehensive theory of reality is based on the reductionist assumption that a theory of particles and forces can explain everything, including life and mind. But all such a theory can do is explain them away. We see that these kinds of "unified theories" do not illuminate the processes underlying emergent phenomena but instead

ignore them as if they were not real—what scientists and philosophers call illusions, or *epiphenomena*—secondary effects or byproducts that have no causal effects on anything.

The reductionist paradigm, at least in its extreme form, has more than a few philosophical implications that most rational people would consider absurd, like the idea that we are not in any sense the makers of any of our decisions or in any way the deciders of our fate. Not only is free will not real to the dogmatic reductionist, neither is basic agency nor the science of biology, nor for that matter, any of the other sciences, like cognitive science, psychology, sociology, or economics. There is only particle physics, and all the rest is simply a story about reality that we've constructed in our minds to make sense of our surroundings for survival purposes. Notice how there's even a paradox lurking within the explanation, which clearly recognizes that we have minds and that there is in fact a real evolutionary process that shapes them for survival. But the whole point of the reductionist philosophy is that the evolutionary process *itself* is epiphenomenal! Genes, too, are just stories we use to make sense of the mechanical trajectories of particles, at least according to this narrative. The reality of the reductionist worldview contradicts the reality we experience at every moment, and for that reason, it leaves all its followers with cognitive dissonance and a need to compartmentalize in the same way that religious fundamentalists need to in order to square their rigid ideology with the facts of life.

We can see the reductionist position clearly in the words of theoretical physicist Brian Greene, string theory's most famous proponent, who said the following to *New Scientist* in an interview promoting his 2020 book, *Until the End of Time: Mind, Matter, and Our Search for Meaning in an Evolving Universe*: "I think it's very important to face up to the truth of reality which is in fact that life and consciousness is a fleeting phenomenon on the entire cosmological timeline . . ."[1]

It's safe to say that Greene does not think that knowledge can make any appreciable difference to the universe at large. But this is not the view of David Deutsch, who writes, "[T]his (evolutionary) process need never come to an end . . . there are no inherent limits to the growth of knowledge and progress."[2]

What is so exciting is that only one of these brilliant physicists can be right. If it turns out that knowledge creation continues forever, or simply *can* continue forever, then a paradigm shift of unparalleled significance is on the horizon. If the universe is coming to life through the spread of adaptive complexity, then nature itself is intrinsically creative.

## Unification over Reduction

Unlike reductionist models, the unifying theory of reality recognizes the higher sciences as legitimate in the sense that they describe real causal phenomena, phenomena that can't be predicted by fundamental physics theories, like quantum field theory (or any other field theory). So, rather than explaining away the higher sciences, by reducing them to the lower ones until they disappear, we have an approach that unifies the disciplines under a common set of laws and principles. These principles describe the emergence and evolution of natural systems in terms of energy and information flows.

Why must any comprehensive theory of reality account for life in its model of cosmic evolution? Because knowledge has extended the causal power of biology on Earth so much that little bits of biosphere—humans in space rockets and man-made satellites—are slowly spreading through the solar system. These are examples of what astrobiologists call *anti-accretion*. While accretion occurs when a planet's gravity pulls in matter from space, anti-accretion is when a planet ejects matter into space. As physicist and origins-of-life researcher Sara Walker explains, anti-accretion typically "requires comprehenders—specifically, the existence of physical systems with *knowledge* of Newton's laws."[3] In other words, we have a cosmic event that can't be explained with any physical theory that does not refer to life, intelligence, and the concept of knowledge.

Alexander Wissner-Gross, American computer scientist and entrepreneur, articulated the same point with a thought experiment that he used to open his 2013 TED Talk on the connection between the second law of thermodynamics and the evolution of intelligence:

> Pretend that you're an alien race that doesn't know anything about Earth biology or Earth neuroscience or Earth intelligence, but you have amazing telescopes, and you are able to watch the earth, and you have amazingly long lives, so you are able to watch the earth over millions, even billions of years. And you observe a really strange effect. You observe that over the course of the millennia, Earth is continually bombarded with asteroids, up until a point. And that at some point, corresponding roughly to our year 2000 AD, asteroids that are on a collision course with the earth, that otherwise would have collided, mysteriously get deflected or they detonate before they can hit the earth. Now of course, as Earthlings, we know the reason would be that we're trying to save ourselves, we're trying to prevent an impact. But if you're an alien race who doesn't know any of this—doesn't have any concept of Earth intelligence—you'd be forced to put together a physical theory that explains how, up until a certain point in time, asteroids that would demolish the surface of a planet mysteriously stop doing that. And so, I claim that this is the same question as understanding the physical nature of intelligence.[4]

*The unifying theory of reality is at once a theory of cosmic evolution and a theory of intelligence*. But to really understand this knowledge-centered scientific paradigm, we must first get a sense of how it is different from the existing paradigm. As explained, the worldview still held by most scientists and crusaders for reason is the reductionist worldview, although that is likely to change soon. This worldview stands in opposition to any sort of *teleological* worldview, which is any cosmic model that gives the universe an intrinsic goal, life an inherent purpose, or evolution an innate arrow or direction.

According to the reductionist philosophy—and it is certainly a philosophy—the universe we inhabit is harsh, chaotic, and entirely impartial to sentient beings, who are merely "happy accidents." In this view, intelligent life is just a temporary statistical fluctuation away from the general trend of increasing disorder. The biosphere is nothing more than a big and complicated dissipative structure, and like all dissipative structures,

destined for transience. Life having some sort of cosmic influence would be a highly unlikely state of affairs, statistically speaking. In the minds of most materialists and reductionists, agency and consciousness can only be epiphenomena, meaning that they have no causal power or cosmic influence.

Interestingly, reductionists looking for existential comfort often try to find significance in life's insignificance. For example, Brian Greene's poetic words are pretty but patently paradoxical:

> [L]ife and consciousness is a fleeting phenomenon on the entire cosmological timeline, which in some sense makes us feel very small, very insignificant, but at the same time, it almost emphasizes what we insignificant beings are able to do; the capacity to tell wondrous stories that help us deal with our own mortality, our own impermanence. It's that dual way of looking at things—we are insignificant but endowed with minds that can reach out to the very edge of the cosmos—that can give a deep sense of purpose and a deep sense of gratitude that we are even here.[5]

This statement is deliciously ironic, because it is precisely the fact that our minds can, as Greene puts it, "reach out to the very edge of the cosmos" that grants life the unique ability to influence its future evolution. It's that "unreasonable effectiveness of mathematics" that Wigner found so mesmerizing. Many decades ago, Freeman Dyson famously wrote, "It is impossible to calculate in detail the long-range future of the universe without including the effects of life and intelligence."[6]

## Life Is Not a Fleeting Moment

The assumption that life is destined for transience and cosmic insignificance initially appears to be supported by one of the most fundamental principles of nature, the second law of thermodynamics. But the narrative that is emerging from complexity science—which unifies physics, biology, and neuroscience by describing them in terms of

thermodynamics and information—tells a different cosmic story. Rather than steadily drifting toward a more disordered and lifeless state, the universe we inhabit is undergoing a transformative evolutionary process that begins before biology and goes far beyond.

The unifying theory of reality recognizes the critical distinction between adaptive complexity and the ordinary dissipative structures found in nature, which spontaneously emerge to consume gradients, then vanish like ghosts, leaving no trace of their existence. As life dissipates the solar gradient, the biosphere evolves, and learns, and grows increasingly robust. For fans of vintage horror flicks, adaptive complexity is the opposite of *The Blob*, which is this amorphous mass of organic gunk that swallows and dissolves all the life around it. Rather than killing beings and converting them into mindless matter, life transforms an inanimate planet into a living one, first by producing biology, and then technology that aids life in its cosmic struggle for existence. As sentient life inevitably emerges, evolves, and expands outward into space, populating planets and saturating inanimate matter with information and intelligence, the cosmos gradually begins to wake up and experience the fruits of its creation.

"Blasphemy!" the staunch reductionist is shouting. That's because most reductionists think that a model of the universe in which life plays some significant role is fundamentally opposed to *naturalism*, the doctrine that says all phenomena must have natural, mechanical explanations. We will see why that is absolutely not the case, but it is easy to see the reason for the confusion.

Since the natural explanations of science are often taken to be the antitheses of the supernatural explanations of religion, it is typically assumed that the position of science should always be in direct opposition to the religious stance, no matter the topic at hand. If religion says that life is special, and that there is some intrinsic purpose to existence, then it is naïvely assumed that science must say the opposite—that life is an insignificant accident in a meaningless and overwhelmingly lifeless universe. We see this mistaken assumption—that a model that grants life cosmic significance is intrinsically religious—in this quote from physicist Sean Carroll from a talk given back in 2015: "You'd expect under theism life to play a special role in the universe; under naturalism you'd expect life

to be very insignificant. I hope I don't need to tell you, life is very insignificant as far as the universe is concerned."[7]

In response to this, I will quote Carroll's colleague, David Deutsch: "There's nothing more deceptive than an obvious fact."[8]

In his book *The Fabric of Reality*, Deutsch suggests that life does indeed play a special role in the universe: "Narrowly conceived evolutionary theory considers us mere 'vehicles' for the replication of our genes or memes; and it refuses to address the question of why evolution has tended to create ever greater adaptive complexity, or the role that such complexity plays in the wider scheme of things."

Freeman Dyson had some words of advice for reductionists who are having trouble coming to terms with the new cosmic narrative presented in this book: "If our analysis of the long-range future leads us to raise questions related to the ultimate meaning and purpose of life, then let us examine these questions boldly and without embarrassment."[9]

## The Universe: A Work in Progress

When people think that intrinsic purpose or progress in nature is a religious idea, it is because they do not have a clear mechanistic understanding of the evolutionary process. They do not see that a Darwinian teleology is not a paradox but the solution to a paradox. You can have inevitable progress without any supernatural or conscious cosmic force guiding or driving the evolutionary process. As will be explained in more depth shortly, a self-correcting system grows more robust and computationally powerful because it is always solving survival problems and storing the solutions in memory. These solutions are structural (morphological), functional (physiological), and computational (behavioral). To use a rough analogy, we can think of adaptive complexity as being similar to a chess-playing AI program that gets defeated again and again, but because it is learning from each failure, it soon becomes unbeatable. As simple as this picture sounds, it just isn't something that has been seriously considered by most physicists or biologists, mostly for ideological reasons.

It is clear that Seth Lloyd, who like Deutsch is a pioneer in the field of quantum computation, does not think organized complexity is transient, and he's not embarrassed to say so. That's because his cosmic model recognizes the causal power of information and the computational nature of our universe:

> Everywhere you look, you see immense variation and complexity. Why? How did the universe get this way? We know from astronomical observation that the initial state of the universe, fourteen billion years ago, was extremely flat, regular, and simple. Similarly, the laws of physics are simple: the known laws of physics could fit on the back of a T-shirt. Simple laws, simple initial state. So where did all of this complexity come from? The laws of physics are silent on this subject. By contrast, the computational theory of the universe has a simple and direct explanation for how and why the universe became complex. The history of the universe in terms of information-processing revolutions, each arising naturally from the previous one, already hints at why a computing universe necessarily gives rise to complexity. In fact, we can prove mathematically that a universe that computes must, with high probability, give rise to a stream of ever-more-complex structures.[10]

From a computational perspective, the growth of adaptive complexity doesn't sound surprising and certainly doesn't seem to violate any sacred scientific principle. But for sociological and ideological reasons, people are more likely to object to the idea of a computational universe inexorably generating "ever-more-complex structures" if the same idea is expressed in casual language. "Life is destined to spread through the universe" just *feels* religious, even though we can show mechanistically why this outcome should be unsurprising, once the larger picture is understood.

American inventor and futurist Ray Kurzweil unabashedly uses "destiny" in describing the future of the universe: "[I]t is clear that the physical laws of our universe are precisely what they need to be to allow for the evolution of increasing levels of order and complexity . . . Ultimately,

the entire universe will become saturated with intelligence. This is the destiny of the universe."[11]

But it is not just technologists, who are known for being kind of out there, trying to get the scientific community at large to realize that *life matters*, and not just to us, but to the cosmos. In his celebrated book *Consciousness: Confessions of a Romantic Reductionist*, neuroscientist Christof Koch, chief scientist and president of the Allen Institute for Brain Science in Seattle, confesses to believing "that the laws of physics overwhelmingly favored the emergence of consciousness." But he doesn't think that complexity growth stops there: "The rise of sentient life within time's wide circuit was inevitable. Teilhard de Chardin is correct in his view that islands within the universe—if not the whole cosmos—are evolving toward ever-greater complexity and self-knowledge."

Knowing that his statement would likely trigger radical reductionists, Koch makes an important caveat that should dissolve the concerns of teleophobic skeptics:

> I am not saying that Earth had to bear life or that bipedal, big-brained primates had to walk the African grasslands. But I do believe that the laws of physics overwhelmingly favored the emergence of consciousness. The universe is a work in progress. Such a belief evokes jeremiads from many biologists and philosophers, but the evidence from cosmology, biology, and history is compelling.

These quotes express the cosmic philosophy of French Jesuit priest and paleontologist Pierre Teilhard de Chardin, who wrote a truly prophetic book about the progressive nature of biological and technological evolution called *The Phenomenon of Man*, published posthumously in 1955, despite being written in the 1930s. Most scholars agree that in it, Teilhard de Chardin predicted the emergence of the internet. While it was heavily criticized by atheists and scientists for its teleological narrative, the Catholic Church considered its evolutionary basis to be heresy, hence the reason for its delayed publication. Similarly influential books were the French philosopher Henri Bergson's *Creative Evolution* (1907), and C. Lloyd Morgan's *Emergent Evolution* (1927). Though they, too, received harsh

criticism from the scientific community upon publication, the emerging paradigm of emergence is rekindling interest in these visionary works.

## Demystifying Evolutionary Progress

We can lessen the allergic reaction to the teleological cosmic narrative by simply referring to life by its more technical name, "adaptive complexity," because the description clarifies why life's inevitable emergence and expansion are mechanistic processes, not mystical ones. Life's mysterious four-billion-year survival streak is no longer so surprising: biology prevails because it is always learning, adapting, and self-correcting. The Darwinian mechanism of variation and selection is a problem-solving algorithm that never fails to find a solution to the existential challenge of survival in a universe that tends toward disorder. Progress occurs because the second law creates a problem that forces its own solution into existence. If an ordered system is to persist, it must adapt. Because life is always copying its form with subtle variations, it does so seamlessly, at first with no conscious effort or awareness.

Chemical, biological, and technological evolution, along with human learning and scientific progress, all involve the accumulation of adaptive information (knowledge) through a deceivingly simple Darwinian mechanism that has been shown to be mathematically equivalent to a process of Bayesian inference. This idea represents the core concept of three scientific philosophies that are actually complementary perspectives of one unified cosmic narrative:

*Universal Darwinism* recognizes that there are evolutionary mechanisms acting at all the different scales in nature, and in doing so, illuminates a continuity between inanimate and animate matter. The entire universe is evolving toward an increasingly complex state, and what is being propagated by life is the knowledge that has accumulated in memory as a result of continual adaptation. It is well established that Dawkins's and Dennett's universal Darwinism, being about a knowledge-creation process, is equivalent to Popper's and Campbell's evolutionary epistemology, which came before it.

*Evolutionary epistemology* emphasizes that the evolutionary process is a learning process, and life is a form of cognition. According to this paradigm, knowledge is created continuously as adaptive complexity evolves and spreads through the cosmos. The universe wakes up not because of some supernatural or conscious teleological force, but because it evolves through a self-correcting process that experiments, removes errors, and encodes survival solutions (adaptations) in memory. The big takeaway is that adaptation is a form of learning, and learning is a form of adaptation.

*Universal Bayesianism* is an update of the two paradigms that came before it. Think of it as "evolutionary epistemology 3.0." This interpretation of the evolutionary process stresses that adaptive systems store a simplified representation or model of the world they inhabit. As evolution and adaptive learning proceed, life is effectively performing statistical inference and updating its model's "beliefs" about the world in a Bayesian fashion. The "free energy principle," the fundamental law of universal Bayesianism, gives us a mathematics, based on statistical mechanics, that describes how life must behave in order to stay in existence. It must minimize its world model's prediction error, and that task entails acquiring knowledge about the universe around it.

While all three philosophies are describing different aspects of the same underlying process, these labels encompass a highly diverse set of phenomena, which are often studied independently and rarely recognized as manifestations of a single universal mechanism that creates knowledge and progress without fail and without limit. We certainly get a picture of cosmic evolution, but it is not perfectly clear from these mechanisms alone that progressive evolution is inevitable and open-ended. To be clear, they explain why complexity is generated, but on their own do not seem to guarantee that the process will continue indefinitely.

By integrating these functionally equivalent theories into a cohesive cosmic narrative, the *evolutionary epistemology-universal Darwinism-universal Bayesianism (EE-UD-UB) framework* offers us a path toward consilience. Scientists have been dreaming of such a consilience since the birth of cybernetics, although one could say it really started with the birth of the scientific method and the ideas of Francis Bacon. It just took about three hundred years for science and math to mature into what was required for a unification of the sciences.

Cybernetics was the first real formal theory of adaptive systems, and from that came the more generalized "general systems theory," a noble attempt to unify the sciences by Ludwig von Bertalanffy, the cyberneticist who was inspired by the work of his colleague W. Ross Ashby, English psychiatrist and cybernetics pioneer, in much the same way Campbell was inspired by Popper.

While general systems theory seemed to have the problem of being too broad, it promoted systems thinking and planted the seed that would develop into modern complexity science. With a focus on complex adaptive systems, complexity science attempts to unify cybernetics and general systems theory with the mathematics of chaos, the statistical mechanics of thermodynamics, and the information theory of Shannon. The latest iteration of this unified systems theory, the evolutionary descendant of all the unifying paradigms that came before it, emerges from the EE-UD-UB framework. I would like to think that Bacon, Bayes, and Bertalanffy would approve of this unifying theory of reality.

## Naturalizing Cosmic Teleology with Natural Selection

It should be stressed that this intrinsically teleological cosmic model does not challenge or refute Darwin's theory of evolution in any way. In fact, it extends the intellectual revolution started by Darwin by explaining the mechanisms of progressive change in the universe on all scales and for all times—past, present, and future—as a result of blind variation and natural selection.

*Cosmic teleology* was never about some mysterious nonphysical force. It was an intuition that life harbored information with cosmic causal power before information had a name. The theory of natural selection doesn't explain away teleology; it naturalizes it and illuminates its mechanistic beauty. The Darwinian dynamic is the driver of evolutionary progress, as suspected by some of the great naturalists of the twentieth century, like de Duve, Dyson, and Morowitz. So, we can no longer conflate the idea of a goal-directed universe with a supernatural or religious theory of existence.

But as we will see in the final chapter, a computational teleology is religious in its own sense. Einstein often said that he was a "deeply religious nonbeliever." This need not be a paradox. Nature's infinite creativity is a source of spiritual inspiration to any observer who realizes they are a manifestation of it.

Unambiguous evidence of continued complexification over time can be found in the same fossil record that confirmed Darwin's great theory. First, unicellular life appeared, then multicellular organisms emerged, followed by mammals, and eventually, humans who could articulate such a process. What's more, the timeline of these events suggests that the rate of progress is accelerating. As the biosphere grows in age, new levels of complexity seem to emerge faster and faster. Zooming out, we see a gradual evolutionary process that is slow to get going but eventually becomes explosive.

The undeniable trend toward higher complexity has led more than a few distinguished scientists to suspect that there may be general laws of nature that drive far-from-equilibrium systems toward ever-greater complexity, just as there are statistical laws that require equilibrium systems to become increasingly disordered. Many are associated with the Santa Fe Institute, like Stuart Kauffman, who suggested many decades ago that something like a "fourth law of thermodynamics" is at work on Earth. In his 2000 book *Investigations*, Kauffman wrote:

> Indeed, there may be a general law for biospheres and perhaps even the universe as a whole along the following lines. A candidate fourth law: As an average trend, biospheres and the universe create novelty and diversity as fast as they can manage to do so without destroying the accumulated propagating organization that is the basis and nexus from which further novelty is discovered and incorporated into the propagating organization.

Ray Kurzweil saw a similar process sticking out of the data like a sore thumb. The "restless genius," as the *Wall Street Journal* dubbed him, does not mince words regarding what he sees as an unassailable truth: "The attributes that are growing exponentially . . . are order and complexity . . . This acceleration matches our commonsense observations."[12]

But what about events that led to mass extinctions? Isn't intelligent life doomed since they, too, seem to be inevitable? What the data reveals is that these mass-extinction events did nothing to buck the overall trend toward greater biosphere diversity and complexity, as evolutionary biologist Anna Dornhaus has pointed out. In fact, they appear to have *increased* the rate of progress. This phenomenon should lead us to wonder whether such catastrophic events are actually a necessary part of an evolutionary process that relies on nature posing life such challenges. Remember Popper's principle from the last chapter: Problems create progress.

But this quasi-optimistic cosmic perspective does not directly follow from the modern synthesis. As traditionally conceived, there is no specific mechanism driving progressive evolution and no apparent mathematical reason to suspect a direction or arrow toward species of higher complexity and intelligence, other than what we'd expect from statistical noise and growing variability around an average. Evolution, as it is understood by most biologists, merely refers to adaptive change, and arbitrary or limited adaptive change does not grant life cosmic significance. For that, evolution must be in some overall sense *progressive*, meaning there must be an arrow or directionality inherent in the process such that life advances toward an increasingly functional and causally effective form.

For this arrow to emerge, the effects of life's adaptations must be cumulative, and complexification and knowledge creation must never stop. Proving that such an arrow exists has been difficult not only because there is no clear, agreed-upon measure of complexity but also no clear definition. This is perhaps the biggest challenge that the unifying theory of reality faces. However, difficult does not mean impossible, and to use the mantra of Deutsch, "Problems are soluble." The problem of quantifying evolutionary progress is no exception. And by solving this problem, we will be making progress.

Of course, to those who hold the reductionist worldview as scripture, some of these new concepts will at first seem blasphemous. While a trend toward greater complexity in the biosphere over time might seem obvious, many evolutionary theorists still reject the notion that replication with natural selection must inevitably lead to the continual emergence of increasingly intelligent forms of life. Although there are some sensible reasons for such objections, which we will now consider, the resistance to

new ideas that support progressive evolution is often motivated by a hidden pressure to conform to an ideological agenda set by atheists bent on protecting science from the corrupting influence of organized religion.

These noble but misguided defenders of logic and reason think that acknowledging evolution to be a cumulative process with an overarching direction amounts to admitting that nature is moving toward some higher goal. And since they by default associate teleology with a supernatural force, rather than a natural thermodynamic, computational, and evolutionary process, many are unwilling to consider any adaptive mechanisms that might contribute to a larger narrative of progress. For this reason, atheists and skeptics will often reject the unfamiliar mechanisms associated with emergence, and in doing so, become science deniers not unlike the religious ideologues they despise.

## The Case Against Evolutionary Progress

One of the most prominent critics of the idea that evolution is creating a "ladder of progress" was the influential twentieth-century evolutionary biologist and popular author Stephen Jay Gould, who argued that natural selection wasn't making organisms more complex per se; it was simply making the whole of life generally more diverse by blindly sending biology down countless evolutionary paths in all directions, seemingly arbitrarily, and with no directed evolution toward some "higher goal." Gould famously described life's helter-skelter exploration of these paths as resembling a drunken sailor's "random walk," which entails stumbling this way or that merely by chance, sometimes veering toward greater complexity, other times away from it. According to Gould, single-celled life like bacteria are the best adapted because they use the least amount of energy—so they have the lowest computational cost—and all other forms of life are just deviations away from that mean. To support this hypothesis, he cited that most organisms on the planet are bacteria, and in the case of a complete cosmic catastrophe, we'd get erased from the face of the earth while they'd probably persist.

While Gould beautifully articulated how the evolutionary process would generate a marvelous variety of species, differing greatly in form and

function, he was quick to emphasize that any appearance of an intrinsic drive toward higher complexity was only a "statistical illusion." Dawkins deconstructed this argument in an otherwise flattering book review,[13] defending the reality of evolutionary progress quite effectively using two main arguments, which we will try to make formal in this chapter and the next. Essentially, evolution toward greater complexity is propelled by an accumulation of adaptive solutions in biological memory (genes, brains, culture, etc.), and also by inevitable major evolutionary transitions, such as the jump from single-celled life to multicellular organisms, and the leap from multicellular organisms to societies, or social organisms.

Gould also passionately preached that human-level intelligence—intelligence that can invent technology and eventually engineer its own evolution—was merely a product of chance and therefore unlikely to emerge again if we could rewind the tape of time and play evolution over. Where Monod, the twentieth-century French biologist mentioned in Chapter One, did not believe that the universe was pregnant with life, Gould did not believe that the biosphere was pregnant with intelligence. To say he loathed the idea that evolution is progressive would be an understatement: "Progress is a noxious, culturally embedded, untestable, nonoperational, intractable idea that must be replaced if we wish to understand the patterns of history."[14]

In science journalist Roger Lewin's wonderful book *Complexity: Life at the Edge of Chaos*, Gould admitted that his beef with the notion of a ladder of progress stemmed from a fear that the philosophy would be used for great evil. And who could blame Gould, who was born in 1941, during World War II, for his concern? Indeed, Adolf Hitler justified his genocidal eugenics program with a moral philosophy based on a distorted version of social Darwinism and a misunderstanding of the concept "survival of the fittest."

In Hitler's interpretation of the notion of a ladder of progress, an obvious bastardization of evolutionary theory, some races were superior to others, even though Darwin went through great trouble to make it known that the "fittest" in this context were not the strongest or even the most intelligent, but the most *adaptable to change*.

Herbert Spencer, a contemporary of Darwin who created the doctrine on social Darwinism, was actually more famous and respected than Darwin at the height of his success, though he fell out of fashion with philosophers even before Hitler inappropriately appropriated his ideas. His belief that society was evolving toward higher complexity was considered teleological and therefore mystical, even though Spencer's model was purely mechanistic, as he was not a religious man.

After World War II, it was pretty much agreed upon by academics that evolutionary thinking should be confined to biology. But as Dennett has stated many times, Darwinian logic is like a "universal acid" in that it cannot be contained: it inevitably spreads out to shape and transform more and more domains.

One could say that Spencer was the first to articulate a theory of universal Darwinism, and in light of the new sciences of complex adaptive systems and non-equilibrium thermodynamics, his ideas were more right than wrong. They do, of course, need to be updated. The new version of Darwin and Spencer's great theory, which includes both competitive and cooperative evolution (phylogenetic and ontogenetic learning), tells us that for fundamental reasons there is *strength in diversity*. This on the surface sounds very different than the notion of survival of the fittest, but we will see how these statements are really two sides of the same coin.

## Making the Case for Evolutionary Progress

Much of what Gould was getting at is true, but that is pretty obvious. As anyone who has taken an evolution course knows, natural selection does not always make an evolving species more complex. If a genetic mutation creates an adaptation that simplifies the design of the creature, and that simplified design increases fitness, then that simpler form will be selected by nature. Critics of the idea of progressive evolution, suspicious of its teleological implications, often cite sharks and crocodiles as examples of species that have not changed in any significant way over many millions of years in existence. Fish that have migrated to caves have been known to lose their eyes over evolutionary time, apparently becoming simpler. This fact clearly illustrates that not all organisms, or even the majority, are

growing more complex through evolution. It would appear that they are simply becoming better adapted to whatever ecological niche they happen to inhabit. Some niches present a great variety of challenges that must be adapted to while others present hardly any. As a result, some organisms barely evolve at all.

As pointed out by American journalist and author Robert Wright, a well-known critic of Gould's position, no one with any sense was claiming that every species was necessarily becoming more complex and intelligent. Cosmic teleologists, religious or secular, do not believe that cockroaches are evolving to become increasingly more intelligent or into something more humanlike. That is the definition of a straw man. But fairness becomes less important, even to scientists, when one believes humanity is at stake. Ironically, Gould's takedown of teleology was inspired by his own teleology, an existential purpose that propelled him to fight against a paradigm that he thought could quell progress.

In summary, we see that like Monod, Gould was both right and wrong. He was right about many of the biological details, but wrong in his interpretation of the larger story implied by those details. While he was correct to point out that natural selection does not drive every species toward increasing complexity, with most species remaining quite simple, he was wrong to conclude from this that evolution is not also continually generating higher levels of computational complexity and intelligence. This unjustified conclusion is common among evolutionary biologists knowingly or unknowingly influenced by Gould's rhetoric and it usually stems from two main mistakes.

The first mistake is ignoring or downplaying the fact that a process that continuously generates a greater diversity of species will also systematically increase the complexity of the most complex species. Once we integrate thermodynamics and information theory into our understanding of Darwinian evolution, we will begin to see why essentially any biosphere being driven away from equilibrium by planetary energy flows for long enough should be expected to produce a species with general intelligence. This is admittedly a bold claim, but one that will prove to be logically unavoidable. To see why, we must understand the cybernetics principle known as the *law of requisite variety* devised by Ashby in the 1960s.

The second mistake is focusing solely on the complexity at the level of the individual organism and not the complexity of the networks formed by these organisms, or the complexity of the biosphere as a whole. In the next chapter, we will try to correct the first mistake. The second mistake, which requires a discussion of the major evolutionary transitions, will be dealt with thoroughly in Chapter Nine.

Efforts to expand the mainstream theory of evolution, known as the *modern synthesis*—an eighty-year-old name it no longer deserves—have collectively created what academics in the field call the *extended evolutionary synthesis*. This update to the theoretical model, a sort of evolutionary theory 2.0, is promoted by a group of respected evolutionary theorists, including philosopher Massimo Pigliucci and evolutionary biologist David Sloan Wilson, who collectively update it as new evolutionary mechanisms come to light.

While it has elucidated a number of overlooked biological mechanisms that inform the larger story—such as the Lamarckian-like mechanisms of epigenetics, evolvability (the evolution of mechanisms of evolution), and the concept of "multilevel selection"—the truth is that it does not go quite far enough. Just as we could not adequately understand how life emerged without thinking about the phenomenon in terms of thermodynamics and information, as suggested by Schrödinger, we cannot really understand how life evolves without doing the same. With this in mind, we will now give our unifying theory of reality a proper name: the *integrated evolutionary synthesis*. Over the course of the rest of the book, this paradigm will bring the concepts from Part One and the previous chapter together to give us a proper "theory of everything"—one that unifies rather than reduces.

# 8

# The Integrated Evolutionary Synthesis

---

> Did Darwin deal a "death blow to Teleology," as [Karl] Marx exclaimed, or did he show how the "rational meaning" of the natural sciences was to be empirically explained (as Marx went right on to exclaim), thereby making a safe home in science for functional or teleological discussion?
>
> **—Daniel Dennett, *Darwin's Dangerous Idea***

---

What is the foundation for a non-reductionist theory of everything? To understand the nature of life, and the nature of reality, we must begin with what many scientists consider to be the law of all physical laws, the second law of thermodynamics. The second law is the basis for our Darwinian teleology.

The natural tendency toward decay and disorder described by Boltzmann's statistical version of the second law—which can be summed up as "things fall apart"—gives any conceivable adaptive or sentient system an eternal existential challenge. Biologists call it survival, cyberneticists call it persistence, and physicists call it staying far from thermodynamic equilibrium. To stay far from equilibrium, in other words, to continue to

exist, a system must be able to extract free energy (what Schrödinger called *negative entropy*) from a noisy, fluctuating, ever-changing environment. This is no trivial task. In fact, it absolutely requires that life *acquire information about the world it inhabits*.

This information has intrinsic meaning to a system because it is causally necessary for its continued existence. Due to its inherent meaningfulness, this kind of information has been called *semantic information, teleosemantic information*, and *adaptive information*, although there are subtle differences between these terms depending on how they are being used. Because adaptive information reduces environmental uncertainty—or Shannon entropy, which is a technical term for ignorance—we also call it *knowledge*. As life fights to maintain its existence, it converts free energy into information, and the cost of this ignorance reduction, this learning process, is the thermal entropy life produces in the form of dissipated heat.

Uncertainty-reducing information is life's first and last weapon in the ongoing war with disorder, and it infuses organic matter with control and causal power. This means that "knowledge is power" is not just a hollow buzz phrase of the digital age; it is true in the most fundamental way. As E. T. Jaynes, the scientist who illuminated the connection between statistical mechanics and information theory, put it: "The old adage 'knowledge is power' . . . is a very cogent truth, both in human relations and in thermodynamics."[1] Without knowledge, life cannot exist for more than a moment, much less colonize the galaxy and beyond. This suggests that setting out on the path toward cosmic superiority is not a choice that intelligent agents like *Homo sapiens* make upon careful reflection, nor is it just some quirky ambition we stumbled onto by chance. Humanity's collective desire to transcend mortality and expand outward into space, so apparent from our scientific and technological endeavors, emerges not incidentally, but as an inevitable consequence of the fact that continual knowledge acquisition is a fundamental biological imperative.

## The Power of Prediction

Like the knowledge we acquire through the practice of science, biological information is based on evidence and predictive, because it is information that allows life to anticipate future events in the world well enough to find energy and avoid threats. It is this predictive power that preserves the complex system, along with the knowledge it embodies, until it can be transmitted to offspring through self-replication. So, not only is evolution equal to learning, evolution and learning are equal to statistical inference, specifically Bayesian inference, or something that approximates it. As evolutionary learning occurs, life is effectively updating its model's "beliefs" about the world around it to be more accurate and comprehensive.

Writing for the publication Aeon, Friston said, "Inference is actually quite close to a theory of everything—including evolution, consciousness, and life itself. It is abduction all the way down."[2] It is the eternal struggle to stay far from equilibrium and the need to anticipate or predict the environment to do so that gives adaptive complexity an intrinsic survival goal and all living creatures an emergent purpose, or teleology.

In 2017, Harvard University cognitive psychologist Steven Pinker answered the question "What scientific term or concept ought to be more widely known?" with the following:

> The Second Law defines the ultimate purpose of life, mind, and human striving: to deploy energy and information to fight back the tide of entropy and carve out refuges of beneficial order. An underappreciation of the inherent tendency toward disorder, and a failure to appreciate the precious niches of order we carve out, are a major source of human folly.[3]

Winning the war against the slow creep of disorder sounds like a challenge that life can or cannot live up to, but when we understand that life is just a vague word for adaptive complexity, we see that it learns spontaneously, at first without any conscious intent to do so. Adaptive complexity begins working on the problem of surviving far from equilibrium—the continual challenge of extracting energy and avoiding threats—the moment it comes into existence. But if that is true, how can life

acquire the information that is causally necessary for its survival before organisms had brains, the organ we associate with learning?

## Phylogenetic and Ontogenetic Learning

As explained in Chapter Six, life begins learning about the regularities in the environment by accumulating knowledge in the genomes of organisms. In this form of generational learning, known as phylogenetic learning, life and death play an equally important role in knowledge creation. Phylogenetic learning is collective learning, but it is competitive rather than cooperative; it is the learning that occurs as a result of the Darwinian dynamic popularly described as "survival of the fittest." This stands in contrast to ontogenetic learning, which is the individual learning we call adaptive learning, or reinforcement learning. If we look close enough, however, individual learning always corresponds to self-organization, so individual learning is actually collective learning at a higher level of resolution! Zoom in far enough and every individual reveals itself to be a community.

An organism is a collection of cells, a brain a collection of neurons, a city a collection of people, a cell a collection of biomolecules—you get the picture. Since all knowledge creation can be understood in terms of either phylogenetic or ontogenetic learning—which correspond to competitive and cooperative evolution—there is a lot to say about each, and even more to say about how they work together. We will focus on the former in this chapter and the latter in the next.

If the second law gives life an existential challenge, Darwinian evolution is how life lives up to the ever-present problem posed by nature. As discussed at length in Chapter Six, the evolutionary mechanism at work whenever there is self-replication with genetic mutation is known as *variation and selection*, and it is a thermodynamic problem-solving algorithm equivalent to trial-and-error learning. It is also functionally equivalent to the mechanism of science, which Popper called "conjecture and refutation," as well as the evolutionary algorithm "generate and test" used in machine learning. Organisms are embodied theories about how to

survive, and science and AI are extensions of life that aid it in its effort to evade entropy.

So, one may think of Darwinian evolution as adaptive complexity blindly exploring the space of possible designs to find solutions to the problem of staying far from equilibrium for the sake of survival. Because the second law's tendency toward decay poses a persistent problem to any self-replicating adaptive system, life will automatically find solutions through blind and mindless experimentation, retaining what works in memory as the errors get filtered out by natural selection. Essentially, mutation is an inventor and natural selection is a pruner. The genetic and neural information that is left over after natural selection culls the generated biodiversity reflects knowledge of the environment that enables the ordered system to survive and thrive. Only the agents with the best designs and most accurate world models get to transmit their information into the future.

## Weeding Out Unstable Energy Flow Channels

Because we are so used to thinking about life from the perspective of the organism, or its genes, we only get glimpses of the true nature of biology. From a planetary perspective, natural selection is a filter that weeds out the unstable energy flow channels, the "errors" at adaptation, and multiplies "solutions," so that functional, well-adapted designs spread through the population. We can understand this as being similar to Shannon's error-correction codes. From an information-theory perspective, natural selection is doing global error correction. As the fittest agents pass the selection filter and self-replicate, the world model of the species gets updated to be more accurate, and that portion of the biosphere becomes increasingly robust.

As external conditions change over time, new adaptive solutions are discovered and written into genetic memory, such that new knowledge is continuously being generated, along with new modes of behavior. This knowledge can be interpreted as information from the outer world transmitted into the interdependent network of complex adaptive systems that we call the biosphere, through the information channel we call natural selection. Once knowledge is transferred from the environment into life,

reproducing organisms propagate this signal, transmitting it to the next generation.

Evolutionary theorists and philosophers of epistemology have been trying to quantify and formalize the evolutionary process almost since information theory was born—first by Fred Dretske, followed by John Maynard Smith, Henry Plotkin, William Harms, Terrence Deacon, and more recently scientists at the Santa Fe Institute like Eric Smith, David Krakauer, and David Wolpert. But what exactly does it mean to say that information is transferred from the world into life?

When adaptations accumulate in an evolving population, the genome of a species is acquiring information about the best strategies for far-from-equilibrium survival and replication. What traits are adaptive will crucially depend on the particular environment the organism inhabits, so if adaptation can be thought of as learning, then it is the niche that provides the education.

As a species adapts to a niche over many generations, the design and behavior of the prototypical organism becomes more statistically correlated with its environment, and therefore it becomes a better predictor of that environment. When the genome of the species accumulates information that reduces the average organism's uncertainty regarding the environment, a model of the environment gets encoded into genetic memory. As the process of natural selection pumps information into the machinery of life, adaptive complexity begins mapping the universe, encoding its structure, and modeling its dynamics. As it does so, the cosmos begins to come to life. Soon after, consciousness emerges, and the universe begins to wake up, bit by bit.

## We Are a Way for the Cosmos to Know Itself

We are the cosmos come alive, not metaphorically, but literally. As theoretical physicist and popular author Max Tegmark likes to point out, sentient beings are particles of inanimate matter rearranged into a configuration that supports computation, causal power, and experience. Looking at the bigger picture, we see that biological evolution, adaptive learning, and scientific progress all reflect an accumulation of uncertainty-

reducing information encoded in genetic, neural, and cultural memory. The integrated evolutionary synthesis—based on the evolutionary epistemology-universal Darwinism-universal Bayesianism framework—recognizes that life, mind, society, culture, science, art, and technology are all manifestations of one evolutionary process, one thermodynamic process, one computational process, unified by the concept of knowledge—adaptive complexity's solution to the eternal problem of uncertainty and disorder.

In his 1995 book *Darwin's Dangerous Idea*, Daniel Dennett wrote:

> If I could give a prize for the single best idea anybody ever had, I'd give it to Darwin. In a single stroke Darwin's theory of evolution by natural selection united the realm of physics and mechanism on one hand with the realm of meaning and purpose on the other. From a Darwinian perspective, the continuity between lifeless matter on the one hand and living things and all their activities and products on the other can be glimpsed in outline and explored in detail, not just the strivings of animals and the efficient designs of plants, but human meanings and purposes: art and science itself, and even morality. When we can see all of our artifacts as fruits on the tree of life, we have achieved a unification of perspective . . .

When one understands that the primary mechanism of cosmic evolution is a problem-solving algorithm that accumulates knowledge in a biosphere as a continual countermeasure to the eternal threat of disorder, a new and radical cosmic narrative begins to emerge whether one is seeking a paradigm shift or not.

We, as sentient beings, truly are engaged in a great cosmic battle, one that quite amusingly turns out to be a story not unlike the eternal wars between good and evil as described by the world's major religions. But the cosmic "struggle for existence," the challenge that faces any sentient species anywhere in the cosmos, is more like the one Deutsch described—a grand and majestic war between order and chaos, life and entropy, existence and nonexistence. Whether one is religious, atheistic, or agnostic, we are all by default engaged in this spiritual war. I mean this in the most literal way; sentience can only continue to exist—feeling and experience and meaning and purpose can only continue to be part of physical reality—if life can

persist in the face of the second law, and the only way it can do that is by acquiring knowledge. This is true whether we are talking about paramecia, plants, poodles, or people.

While our individual success, or the success of our civilization, is in no way guaranteed, it appears that the game is rigged in life's favor, thanks to the miracle of self-correction. As adaptive complexity grows increasingly diverse and complex over evolutionary time, it becomes harder to kill or restrain. This is the magic of the error-correcting mechanism. By learning from its mistakes, the biosphere becomes more powerful from everything that doesn't destroy it completely. Indeed, what doesn't kill life makes it stronger.

So, the inevitable growth of knowledge and spread of adaptive complexity in the cosmos is not driven by some supernatural or conscious force; it is a learning process that creates knowledge that keeps life out of thermodynamic equilibrium, and it arises from an evolutionary process that is equivalent to Bayesian inference. This equivalence has been demonstrated by John Campbell and Karl Friston.

The integrated evolutionary synthesis extends Darwinian evolution beyond the biological domain, and it illuminates the evolutionary mechanisms acting at all the different scales in nature, which constitute both the hierarchy of matter and the hierarchy of life. We will talk about these hierarchies in the next chapter, which is about emergence. All this certainly gives us a picture of cosmic evolution, but again, it is not clear from these mechanisms alone that the evolutionary progress on Earth is part of an inevitable cosmic process. For that, we need the cybernetics principle known as the law of requisite variety, which will give us a new perspective on the second law of thermodynamics that suggests evolution is not just progressive but inevitably leads to an intelligence far greater than anything we've yet to imagine (or can imagine).

## Evolution's Arrow Emerges from Knowledge Accumulation

Stephen Jay Gould's view on evolution tells us that there is a longstanding debate about whether natural selection is or is not making life increasingly complex. As we've established, the answer to that question is yes and no, for perfectly good reasons.

A well-adapted species represents an incredibly resilient but often temporary biological solution to an existential thermodynamic dilemma—a sort of living, evolving scientific theory about how to most efficiently extract the free energy, the lifeblood of existence, out of a particular niche. And just like scientific theories conform to Occam's razor (in finding the simplest way to solve problems), so do species. Einstein famously said, "A theory should be as simple as possible, but no simpler." The same could be said for an organism.

Naturally, the complexity of the organism will evolve to match the complexity of the niche. More specifically, the complexity of the organism will correspond to the complexity of the environmental challenge the organism is confronted with, which starts off as a purely thermodynamic challenge. Some niches present a changing variety of challenges that must be adapted to while others present hardly any. As a result, some species don't need to evolve. Others, however, have niches that become increasingly complex over time.

Taking the thermodynamic perspective of evolution, we can think of each niche on Earth as a sort of "energy slot" for a given species, and an evolving population of organisms efficiently searching the solution space will eventually, just by chance, discover a solution to a thermodynamic problem that it didn't know existed. The discovery of a novel energy source or energy-extraction technique is both how a new niche and a new species come into existence. Phylogenetic learning will naturally lead to speciation because organisms will stumble upon ways to exploit thermodynamic niches that were previously inaccessible to life purely for design reasons. Olivia Judson, an evolutionary biologist and a science writer, describes what she calls "the energy expansions of evolution" in a 2017 paper published in the online journal *Nature Ecology and Evolution*:

> The history of the life-Earth system can be divided into five "energetic" epochs, each featuring the evolution of life forms that can exploit a new source of energy. These sources are: geochemical

> energy, sunlight, oxygen, flesh and fire. The first two were present at the start, but oxygen, flesh and fire are all consequences of evolutionary events. Since no category of energy source has disappeared, this has, over time, resulted in an expanding realm of the sources of energy available to living organisms and a concomitant increase in the diversity and complexity of ecosystems.

So, as the biosphere accumulates knowledge through phylogenetic learning, not only is a solution space corresponding to one kind of thermodynamic problem being explored, there also is a growing "problem space," each with its own unique "solution space." This growing design space would seem to exist in a platonic sense from life's beginning, a future garden of biodiversity just waiting to be discovered. The evolution of the biosphere is not perfectly predictable or determined in any strict sense of the word, as contingency ("accident") is part of how evolution works, but it is determined in a statistical sense, in terms of a predictable distribution of far-from-equilibrium attractors of varying degrees of complexity, which inexorably emerge from a thermodynamic planetary process.

The first life-forms, now thought to be reductive autotrophs, were the solution to the problem of how to extract free energy from the geochemistry supplied by hydrothermal vents. They had a narrow task, but replication with mutation eventually produced innovations in energy-extraction technology that led to many different species of bacteria.

After all of these kinds of niches became occupied, there was still an ocean of untapped free energy striking Earth from above. Photosynthetic bacteria, the single-celled ancestors of plants, were a solution to the problem of how to extract work from all the solar energy that was flowing through the planetary system. And heterotrophic organisms—organisms that eat other organisms to survive—were a solution to the problem of how to extract energy from life itself. Once life starts having to model life—especially other agents with causal power and adaptive behavior—the computational task of extracting free energy gets increasingly difficult. As more complex forms of life emerge and become potential prey for predators, food sources become harder to predict, and more sophisticated world models are needed.

By virtue of having to model each other, the complexity of carnivorous species gets ratcheted up by what is known as an "evolutionary arms race." The gazelles that have a new way to outsmart lions get to reproduce, and the lions who are clever enough to anticipate the gazelle's new trick get to reproduce. For some species, simply persisting requires a continual increase in computational sophistication. This concept is embodied in the *Red Queen hypothesis*, well known to evolutionary theorists and cyberneticists.

The theory proposes that species must constantly adapt, evolve, and reproduce just to maintain their existence, due to a competitive, evolving environment. The name, proposed by the evolutionary biologist Leigh Van Valen in 1973, comes from Lewis Carroll's *Alice's Adventures in Wonderland*. After Alice complains of running for a very long time and going nowhere, the Red Queen responds, "Now, here, you see, it takes all the running you can do to keep in the same place." In other words, some species must continuously evolve toward higher complexity just to stay in the game of existence.

The integrated evolutionary synthesis tells us the general diversity of species we see in the biosphere comes from mutations that discover solutions to new thermodynamic problems. Because each ecological niche has different physical characteristics and variables, life diversifies out of necessity. An evolving species is learning about the universe, even if it only models a very small slice of it. With millions of different species alive today, that's a whole lot of knowledge about how biology can survive and thrive in various real-world conditions.

## The Laws of Life

Life doesn't want to overthink the problem of staying far from equilibrium, because thinking is computation and information processing costs energy, energy that could be used to stay far from equilibrium for longer, which would allow for more self-replication. Biology is always trying to use energy more efficiently so that it can squeeze more work out of the same resource supply. This means that as life evolves, it becomes not just better at capturing energy, but more energy efficient.

The thermodynamic cost of all the computations life performs to stay out of equilibrium has been calculated by David Wolpert, and this cost is close to the *Landauer limit*, which represents the most efficient computation possible in terms of energy demands. Evolving to do more with less is known as *ephemeralization*, a term coined by futurist Buckminster Fuller that's really nothing more than thermodynamic optimization. Fuller described ephemeralization as the ability of technological advancement to "do more and more with less and less until eventually you can do everything with nothing." We may not be aware of it, but technological progress, like that which drives Moore's Law, is about making computation more powerful by making it more energy efficient. The same could be said for evolution more broadly.

The integrated evolutionary synthesis gets rid of the fitness tautology: the problem of fitness defined as that which aids survival, and that which aids survival being defined as that which is fit. From the thermodynamic perspective, fitness pertains to how well an organism can resist entropic decay. In particular, fitness corresponds to the resilience of the embodied energy-extraction program, the dissipative channel, and is not correlated with strength, intelligence, or even complexity. Being fit in an evolutionary sense does not mean being the strongest, smartest, or fastest; a sloth is none of those things, but it has survived for many millions of years despite its general slothyness. Fitness is about being able to effectively exploit a source of environmental energy and avoid threats that can puncture the Markov blanket and harm the integrity of the ordered system.

Logically, we can state that mutations that inhibit an organism's ability to meet its energy requirements will be selected against, while mutations that improve the ability to utilize energy will be selected for. This can be summed up by what is known as the *constructal law* developed by Adrian Bejan, a Duke University professor whose work has done much to advance our understanding of thermodynamics. In Bejan's words, the law states: "For a finite-size flow system to persist in time [to live], its configuration must evolve in such a way that provides easier access to the currents flowing through it."[4]

If one thinks about it, this law is a restatement of Morowitz's 1968 quote, "The energy that flows through a system acts to organize that

system," introduced in Chapter Four. Energy flowing through chemical networks forces intricate patterns of biological design into existence as if the universe were a big factory built to do so. This law has been formalized to some degree, albeit abstractly, in England's work on dissipative adaptation, and also in Bejan's own work, which applies the constructal law to all kinds of systems—chemical, biological, social, economic, even technological.

Seven years before the release of the 2012 book Bejan coauthored, *Design in Nature*, similar ideas were presented in the popular science book *Into the Cool: Energy Flow, Thermodynamics, and Life*, written by geologist Eric Schneider and Dorion Sagan, son of Carl Sagan and Lynn Margulis. Though Sagan and Schneider rightly recognize the role life and civilization play in opening up channels for energy flow, they appear to have drawn the wrong philosophical conclusions about the nature of the thermodynamic process. What they fail to see, which Morowitz clearly recognized and articulated in his 2002 book *The Emergence of Everything*, is that producing entropy is not "life's purpose," but instead a byproduct of computation, complexity increase, and knowledge creation. It is the energetic cost of cosmic self-organization.

Producing entropy is life's purpose no more than your computer's purpose is to produce heat or your car's purpose is to produce exhaust. Life opens up energy flows that organize matter into computational machinery that can model, navigate, and understand the world around it, and the cost of that computation is some usable energy being converted into unusable energy, or entropy. Computation, which is mechanical work, has a price. That is the simple idea underlying the second law of thermodynamics as formulated by Carnot and Clausius, but it was impossible to see clearly before we had a proper theory of information and computation. First, Rolf Landauer, a lauded researcher at IBM, had to show why "information is physical." Bits of information must be stored on bits of matter. The idea that information is physical doesn't sound too surprising now, but in the 1950s and '60s, it was radical.

## Redefining Fitness in Thermodynamic Terms

Once we understand that life's thermodynamic imperative necessitates knowledge accumulation, we see that natural selection is maximizing fitness by improving the robustness of the energy flow channel and optimizing the accuracy of the organism's predictive model. These might sound like two different functions on the surface, but they are really descriptions of the same process. The ability of the organism to perform its dissipative function is determined by how well the agent can extract free energy, which is dependent upon how well its internal model predicts its environment.

We must not forget that the internal model must be *embodied* in a system with a mechanical design that allows the agent to benefit from the accuracy of the model. But, again, the model and the embodiment of the model are really describing a single design, because the predictive model is the evolutionary result of an accumulation of physical adaptations that allowed for more efficient energy extraction. It's been said that evolution is an engineer, but it is also a computer modeler.

Keeping in line with past nomenclature, we can define fitness a bit less tautologically and a bit more precisely as *dissipative fitness*, or how well an adaptive system can stay far from equilibrium by extracting free energy. This energy-focused characterization of fitness does not require any changes to the formalisms evolutionary theorists have developed, only a shift in perspective motivated by a fundamental reconceptualization of life as an energy-dependent, steady-state, non-equilibrium system that must acquire information to reduce environmental uncertainty sufficient to stay far from thermodynamic equilibrium.

To understand what we mean when we say that life performs computation near optimal thermodynamic efficiency—to know what getting close to the Landauer limit looks like—let's consider a familiar example. It is well known that the human brain is constantly performing tasks that our best supercomputers and autonomous robots fail at miserably, and it does so with only the amount of energy it takes to run just one household lightbulb.

This suggests that as a result of natural selection, individual agents minimize rather than maximize the rate of energy dissipation or entropy production, as Prigogine noted many decades ago. Your literal purpose in life is not to create as much disorder (entropy) as possible—as some have interpreted the thermodynamic narrative—but to acquire knowledge that

preserves and expands the order constructed by life. However, the collective result of agents who behave in this way is a biosphere that grows increasingly diverse, complex, and computationally powerful over time, such that the rate of total entropy production is maximized over a time window that extends into the future.

That is to say, the universe that produces entropy at the fastest possible rate over its lifetime is the universe in which life emerges and spreads at the fastest possible rate. Wissner-Gross describes this fundamental connection between the laws of physics and the increase in cosmic complexity in the same TED Talk mentioned in the previous chapter:

> I looked at a variety of different threads, across science, across a variety of disciplines, that were pointing, I think, towards a single underlying mechanism for intelligence. For cosmology, for example, there have been a variety of different threads of evidence that our universe appears to be finely tuned for the development of intelligence, and in particular for universal states that maximize the diversity of possible futures.

But what does it mean to say that the universe is tuned for the emergence of universal states that "maximize the diversity of possible futures"? This is undoubtedly an unfamiliar notion of intelligence, but we will see why it is the most straightforward way to understand what we are talking about when we use those words. Stated simply, life does not like to get trapped. Intelligent agents seek freedom, and that freedom comes from an ability to respond adaptively to a greater diversity of challenges.

## Evolution Increases Empowerment

Wissner-Gross's theory has an older analog in the computational neuroscience and machine-learning literature called *empowerment*, proposed by UK artificial intelligence researcher Daniel Polani and colleagues. Both theories of intelligence are being used to create more intelligent-behaving artificial systems. Instead of teaching robots to perform

individual tasks, they are programmed to follow a more general rule, one of an existential and practical nature: seek to keep future options open.

"Empowerment means being in a state where you have the greatest potential influence on the world you can perceive," said Christoph Salge, who developed the idea with Polani, in a press release. "So, for a simple robot, this might be getting safely back to its power station, and not getting stuck, which would limit its options for movement. For a more futuristic, human-like robot this would not just include movement, but could incorporate a variety of parameters, resulting in more human-like drives."[5]

For humans, "maximally empowered" might mean owning a car, having wealth, or being an influential person (or simply having influential friends or family members). Of course, one needs be none of these things to be optimally empowered given a set of goals in life, but even less ambitious humans do not want to be jailed or injured in a way that makes surviving or adapting to new circumstances difficult. We all want to keep our future options open to some degree, and that means simultaneously having stability and flexibility in responding to challenging situations. Maximizing future freedom of action is all part of the general effort to stay far from thermodynamic equilibrium, so it is fitting that Wissner-Gross has called this natural tendency a "causal entropic force."

If causal entropic forcing is real and connected to the second law of thermodynamics, maybe we don't need a fourth law. Maybe we just need a reinterpretation of the second law that explains its connection to the emergence and evolution of intelligence. Wissner-Gross has publicly stated that he is actively searching for such an explanation. By turning to the cybernetics principles that inspired ideas such as empowerment, we will arrive at an evolutionary explanation for why causal entropic forcing inevitably produces intelligent agents of increasing computational power.

## The Relationship Between Entropy and Uncertainty

Remember our discussions of entropy in Chapter Two, and recall the importance of the microstate-macrostate distinction. A macrostate would be

something like an average temperature of a system of gas molecules (a global measure), while the microstate is a description of the position and velocity of every molecule in the system (which turns out to be impossible to measure). From clever experiments, we have confirmed Boltzmann's intuition that there are many different microstates (specific particle arrangements) that correspond to a single macrostate. This is because there are many equivalent ways for the individual molecules to be spread out and have the same average total kinetic energy.

What Boltzmann showed was that entropy is a measure of how many different ways a system can be arranged without changing its macrostate. The higher the entropy, the more microstates there are that correspond to a single macrostate. The reason systems at or near equilibrium naturally drift toward higher entropy or disorder is because there are simply many more ways to be equivalently "mixed up" and disorderly than there are ways to be organized and patterned. Systems naturally drift toward disorder from the large-scale effects of chance.

But with open systems, that random drift gets biased by a flow of energy that pushes the system toward an attractor, constraining chance and transforming it into necessity or inevitability. The system can no longer be described as *ergodic*, meaning the system does not explore the configuration space randomly and exhaustively; non-ergodic systems, such as adaptive systems, evolve toward configurations that save the system from dissipating into nonexistence. The non-equilibrium attractor the system moves toward represents a configuration that is engineered for environmental energy extraction. Self-organization is what we call a system moving toward an attractor, and it organizes itself through the Darwinian dynamic called *blind variation and selective retention*. To sum up the natural selection of self-organizing systems simply: What works, persists.

In the 1960s, E. T. Jaynes showed that Boltzmann's entropy is not just a measure of disorder or the number of microstates that correspond to a particular macrostate. Entropy is also a measure of the uncertainty or ignorance of an observer observing the system's macrostate and not knowing the specific microstate. Because states with higher disorder have more equivalent microstates that correspond to a unique macrostate, higher entropy means greater ignorance—and less certainty—about the specific microstate the system is in. It also means there is more surprise, and

information gained, upon learning the actual microstate of the system, because more is learned (more bits of information gained) when there are more alternative possibilities that get eliminated by the act of measurement and observation.

Of course, as we've established, perfectly measuring a microstate is impossible because of quantum uncertainty and classical-scale chaos, but this narrative allows us to think about entropy in terms of information, and in doing so, we see that entropy is curiously related to the knowledge of the observer, and the accuracy of their model of the system under observation. This is why statistical thermodynamics and Shannon's information theory are essentially the same theory: Shannon entropy is a measure of how many states a system can be in, or how many alternative messages can be sent over a channel. We discussed this at length in Chapter Five. The big thing we need to know is that entropy can be a measure of the number of states a system can be in or the number of ways it can be configured. For this reason, entropy measures can be used to measure complexity, but to really understand why, we must briefly visit two cybernetics principles: Ashby's good regulator theorem and law of requisite variety.

## The Good Regulator Theorem and the Law of Requisite Variety

The *good regulator theorem* was introduced by W. Ross Ashby and Roger Conant in an influential 1970 paper titled "Every Good Regulator of a System Must Be a Model of That System." Six years later, the theorem would be repackaged as the internal model principle, which is a critical component of control theory. The good regulator theorem—as applied to evolutionary biology—states that any adaptive system (organism) must model the relevant environmental variables to continue persisting. Relevant variables are any that pertain to the intrinsic thermodynamic challenge of staying far from equilibrium. Energy extraction, threat avoidance, and remaining ordered in the face of environmental disturbances are critical efforts that permit a sentient system to stay in the game of existence, to persist for a little while longer.

While we can reason that a bacterium performing chemotaxis (movement in response to a chemical stimulus) and a plant performing heliotropism (moving toward the direction of the sun) have modeled their source of energy in the most general sense of the word, these theorems give us a mathematical formulation for why that must be true. There's really no way to understand an organism without understanding its relationship to the environment because—as Friston reminds us—the organism itself is a model of its environment. There may be no subjective experience associated with the world model encoded in a brainless organism, but it does have some sort of abstract map of the territory that gives the system a statistical understanding of what it can expect to encounter.

Ashby's *law of requisite variety* is a principle from cybernetics that tells you something about the computational sophistication of an agent's internal model. This law—as applied to evolutionary biology—says that to continue to survive, an organism must have a repertoire of states that is at least equal to the number of different challenges or disturbances presented by its environment, and it gives us a mathematical explanation of why this principle must also be true.

John Naughton, a senior research fellow at Cambridge University, sums up the law this way:

> In colloquial terms Ashby's Law has come to be understood as a simple proposition: if a system is to be able to deal successfully with the diversity of challenges that its environment produces, then it needs to have a repertoire of responses which is (at least) as nuanced as the problems thrown up by the environment. So a viable system is one that can handle the variability of its environment. Or, as Ashby put it, only variety can absorb variety.[6]

Think about it like this. A cat who eats mice to survive must have at least as many behavioral states as ways the mouse can evade it. Conversely, the mouse has to have enough behavioral states to get away. In the same way, a swordsman must have as many blocks as his opponent has attacks. Each system has a model of the other. And, because all the various behavioral responses map onto unique internal states, the law says that an organism must have as many cognitive states as is required by the complexity of the

niche, which is determined by the number of different thermodynamic challenges the environment poses to the organism.

So, the law of requisite variety suggests that a species' intelligence corresponds to the number of accessible states in its behavioral or mental repertoire—which won't be the exact same size, but for now this is a helpful simplification. To quantify this number, we can describe it in terms of entropy: the more possible states the cognitive system can be in, the higher the entropy, and presumably, the higher the intelligence. In this application, entropy is not a measure of disorder but a measure of cognitive bandwidth. It is also a measure of our ignorance of the exact mental state of an organism at an instant in time if we cannot observe the organism's state directly. Just like in Jaynes's statistical mechanics, entropy is a measure of the ignorance one has when one does not know the precise state of a system under observation, which in this example would be a specific neural state instead of a microstate.

Neuroscientist Robin Carhart-Harris of Imperial College London, whose entropic brain hypothesis attempts to explain the effects of psychedelics on states of consciousness, cites integrated information theory's *phi* (we will come back to this term later) as a measure of information entropy for brains, although the same theory can be applied to brainless organisms as well: "The view taken here is that the human brain exhibits greater entropy than other members of the animal kingdom, which is equivalent to saying that the human mind possesses a greater repertoire of potential mental states than lower animals."[7]

Christof Koch, perhaps the most famous proponent of integrated information theory, said, "The more complex the system, the larger the repertoire of conscious states it can experience."[8] By quantifying the number of cognitive states that can potentially be accessed by an organism, entropy measures can be used to evaluate the computational complexity of any sentient system.

## Niche Emergence Drives Complexification

So why has the evolutionary record shown a trend toward increasingly complex forms, in terms of how many states the system can support, if species are only evolving to be as complex as some finitely complex niche? Why should the process produce increasingly intelligent agents?

Two words: *niche emergence*. Every new species that emerges creates a new niche by serving as a potential food source for a new species. Continual niche emergence leads to increasingly complex species because the most complex species will have to model an increasingly complex-behaving source of energy and an increasingly complex external world. In fact, humans in modern society, in the "age of information," must model other modelers, sometimes in detail. Think about the job of a psychologist or a competitive athlete like a boxer. So, at the present moment in history, adaptive systems actually model more than just the external environment; they model other minds.

As E.O. Wilson explained in his 1992 book *The Diversity of Life*, there is a self-reinforcing tendency for ecosystems to create new niches and new species continually. Because the number of biologically relevant variables grows open-endedly as this process continues, the law of requisite variety says that the most complex species' world model must become increasingly complex also. If intelligence is correlated with the number of internal states that are accessible to an organism, which can be measured with phi, it should not be surprising that increasingly intelligent species emerge over time as a result of evolution. As the number of challenges posed by an environment grows, so does the size of the mental repertoire that is required to respond to those challenges.

As discussed in the next chapter, an ecosystem is functionally equivalent to a chemical autocatalytic set, which actively reproduces the members of the set. It also grows the set to include new members as it continuously adapts to the energy landscape through the positive feedback loop we call *self-amplification*. Of course, a species cannot model an infinite number of variables, so the optimal predictive model would be one that at some point adopts a general intelligence strategy. Now the brain doesn't have to model every single new variable that could or could not be food, that could or could not be a threat, in precise detail.

Francis Heylighen, a Belgian cyberneticist who has written about the connection between the law of requisite variety and the second law of

thermodynamics for decades, said Ashby's law should be of interest to any society that is hoping to avoid existential threats like natural and man-made disasters: "This principle has important implications for practical situations: since the variety of perturbations a system can potentially be confronted with is unlimited, we should always try to maximize its internal variety (or diversity), so as to be optimally prepared for any foreseeable or unforeseeable contingency."[9]

Although it is easy to see how selection pressures eventually produce an intelligent species, Gould is right that there is a limit to how intelligent a species can become as a result of biological evolution alone. But with the integrated evolutionary synthesis, we see that our natural evolution includes merging with our technology. This means that if our aim is to measure the intelligence of life, we must also measure the computational capacity of the biosphere as a whole, which includes the layer of technology that sits on top of it.

The growing ladder of increasingly complex niches creates a distribution of interacting species that vary in shape and size but fit together to form a greater functional entity—a distributed intelligence that stretches across the planet. This global cybernetic system has been named "Gaia" for reasons that will be explained in the next chapter, and it is also "trying" to maximize the number of environmental states it can respond to by increasing its internal variety, or the number of accessible computational states. As a complex adaptive system, the biosphere behaves like an autopoietic agent, and our technological progress is a manifestation of Gaia's attempt to evade equilibrium indefinitely. Without technology, life cannot leave the planet, which it must do before the sun explodes. For life to do its dissipative duty, intelligence must persist.

Now we can see the connection between the second law of thermodynamics and the law of requisite variety. Adaptive complexity must continually expand the size of its repertoire of cognitive states to deal with the growing list of thermodynamic challenges it faces (those problems that threaten its far-from-equilibrium existence). By growing more intelligent, the biosphere, an interconnected, interdependent, integrated computational network, dissipates the solar-induced gradient more efficiently, producing thermal entropy at the fastest rate allowed by the laws of physics. But, at

this point, we should no longer think of the universe as trying to produce entropy at the fastest rate possible; the universe is self-organizing at the fastest rate possible, and the increase in entropy is just the energetic cost of the work required to construct cosmic order.

## Energy Flow as a Measure of Complexity

In his 2001 book *Cosmic Evolution: The Rise of Complexity in Nature*, Harvard University astrophysicist Eric Chaisson proposed a quantitative measure of complexity called the *energy rate density*, which is the energy flow per second and per gram of matter. The idea is that more complex machinery will require more energy flow and will produce more thermal entropy. According to this metric, planets are more complex than stars, but not as complex as plants, which are less complex than animals. More complex than animals are societies and technologically advanced civilizations. Presumably, the information processing occurring in these systems is the source of energy consumption.

Chaisson has also attributed the inevitable increase in complexity to Darwinian evolution, but with a thermodynamic spin that reinforces the narrative that has been developed in this part of the book: "Natural selection exploits energy flows, determining which flows are conducive to the system, thereby apparently optimizing them."[10] While the rate of energy flow sounds like a promising quantitative measure of complexity, it doesn't seem to tell us much about the nature of intelligence, other than it guzzles up free energy. Entropy production must correlate in some tight way with information-processing capacity.

In his 1992 book *Complexity: Life at the Edge of Chaos*, British science writer Roger Lewin quoted American complexity scientist and chaos theorist Norman Packard: "Biological complexity has to do with the ability to process information . . . Computational capability, that's what we see in our cellular automata models, and in other complex adaptive systems, and what drives their evolution is increased computational ability."

Astrobiologist Sara Walker made the same point as Packard in her award-winning essay "Bio from Bit": "There is a lot more information in

the biosphere today than there was at the time of the last universal common ancestor ~3.5 billion years ago. Presumably, this trend is driven by evolution encoding more information in biological networks over time through selection."

Wissner-Gross's causal entropic force is the engine for the open-ended Darwinian exploration of life's design space—what Stuart Kauffman has called the "adjacent possible." Life evolves to maximize empowerment, which means maximizing future possibilities and preserving future freedom of action. In doing so, cybernetic systems carve out a cosmic trajectory that was inaccessible to inanimate forms of matter. Life does not violate the laws of physics in any way, but it does create new ones—emergent laws and dynamics that require computational and statistical description. Scientists, however, typically ignore what they can't measure. E.O. Wilson once said, "It is not difficult to recognize complexity . . . The difficulty comes in how you measure it."[11]

We are now equipped with the concepts required to define and measure the kind of complexity that corresponds to intelligence.

## A New Measure of Complexity

When dealing with something complicated and hard to pin down, it often helps to keep things simple. Following the logic made popular by John Maynard Smith, an aeronautical engineer turned evolutionary theorist, systems with a greater variety of parts and a greater number of connections between those parts should typically be considered more complex. Having more connections means the system is more integrated, and more variety or distinction among parts means the system is more differentiated.

Complexity would then correspond to a single number that would represent the degree to which a complex system was both differentiated and integrated, and presumably that number would represent the size of the repertoire of accessible internal states. We already know we have that measure, and it is integrated information theory's phi, which extends Shannon's information theory to apply to adaptive systems using

formalisms from Israeli American computer scientist and philosopher Judea Pearl's causal calculus.

Phi is an information measure that reflects both differentiation and integration in a single metric, and it is thought to be correlated with a system's intelligence, and some believe its "degree of consciousness." While talking about consciousness can get tricky, phi provides a precise mathematical measure of computational complexity that corresponds to (1) the system's cause-effect power over itself and on its environment, and (2) the number of cognitive states it can potentially access. Obviously, this is a massive number to try and measure, but it's no worse than the practical difficulty Boltzmann's great theory of statistical mechanics faced, as the numbers are similarly enormous when we talk about microstates of a system. What we do to make things easier is invent approximate or "proxy" measures for phi, which neuroscientists have already done with clinical success, as we will see in Part Three.

Koch further describes phi as a measure of *synergy*, defined as the degree to which a system is "more than the sum of its parts." This suggests phi can be used to identify examples of emergence in nature. As he and Carhart-Harris have pointed out, phi is correlated with the size of the repertoire of accessible mental states, such that higher phi systems have a more diverse set of responses and are thus able to handle more complex environments. You could say that phi is a measure of a system's requisite variety. Now we have made a mechanistic argument that unites the second law with the law of requisite variety to produce a principle of increasing cosmic complexity, which inevitably generates increasingly empowered adaptive agents.

If you find this new model of cosmic evolution to be unacceptable because of the central role that life plays in the universe's future development, remember that it is entirely mechanistic. We cannot ignore the progressive nature of evolution just because it contradicts our reductionist assumptions. We must embrace life's cosmic significance—and all the spiritual implications that follow from this fact. One such implication is that we are indeed "special," despite being a product of Darwinian evolution (rather than a supernatural force), as John Campbell reassures us in his book *Universal Darwinism: The Path of Knowledge*: "This view of Universal Darwinism provides us with a unified and integrated view of the evolution

of information processes and knowledge entities. Although this view is scientific, it does acknowledge that humans have a special place, as growing tips, on this ageless, sprouting tree of knowledge."

We are now familiar with phylogenetic learning and the basics of ontogenetic learning, or self-organization. In the next chapter we will see how both mechanisms work together to produce hierarchical control systems of increasing computational complexity and causal power through a process of recursive emergence.

# 9

# Hierarchical Emergence

---

> There are things Darwin couldn't have known. One of them was self-organization in complex dynamical systems. If the new science of complexity succeeds, it will broker a marriage between self-organization and selection. It'll be a physics of biology.
>
> **—Stuart Kauffman**

---

It is often assumed that life is cosmically insignificant because it occupies such a small region of the universe, a universe that seems so hostile and inhospitable to life. But both of these assessments are shortsighted. If the unifying theory of reality presented in this book is correct, it would suggest that life will inevitably spread through the cosmos at large, dominating the universe in very much the same way that life dominates planet Earth today. In this chapter, we will build on the idea that life perseveres not *in spite* of the existential challenges the universe presents it with, but *because of them*. Popper's principle, *problems create progress*, explains why life naturally grows more complex and powerful as a result of evolution. Simply stated, existential challenges force life to continuously seek out survival solutions.

These solutions, which we call adaptations, are incremental improvements on life's design.

Life has the eternal problem of remaining far from thermodynamic equilibrium, and it achieves this goal by acquiring information about how to survive in its environment. How is this adaptive information acquired? Through blind variation and natural selection, which is a form of trial-and-error learning, as evolutionary epistemology and universal Darwinism so clearly reveal. In this way, the growth of knowledge is both a Darwinian and a thermodynamic process. Let's make this connection clear by summarizing the physical story as it has been told so far. Doing so will lead us to a new natural philosophy that explains existence in terms of evolution and emergence.

## A New Spin on the Second Law

From the thermodynamic perspective established in Part One, we know that biology's function is to provide energy flow channels for a chemically stressed planet, because in doing so, life facilitates entropy production. We call these energy flow channels organisms, and they come in many different varieties or species. Since the influx of solar energy is continuously driving the planet away from equilibrium, entropy maximization requires that those energy flow channels persist into the future. From this perspective, biological evolution is a process that builds and maintains energy flow channels to facilitate the dissipation of planetary free energy. To perform its thermodynamic function, life must acquire information that reduces environmental uncertainty, so adaptation is a process of statistical modeling that is equivalent to Bayesian inference. As life evolves and learns and updates its model's "beliefs" about the environment to be increasingly accurate, it minimizes its ignorance regarding all the possible ways the surrounding world can surprise it. Science is just an extension of our intrinsic need to predict the universe so that we may stay in the game of existence.

As evolution proceeds in this fashion, nature builds a diverse array of adaptive systems that model the world around them in an effort to evade equilibrium. Because the emergence of increasingly complex environments

is inevitable, so, too, is the emergence of increasingly intelligent species. This brings us to a cosmic conclusion that is very much the opposite of the initial interpretation of the second law of thermodynamics. To avoid confusion, maybe we should follow Stuart Kauffman's advice and just call it the *fourth law of thermodynamics*. Simply stated, the fourth law means: *A universe that maximizes entropy, or minimizes free energy, at the fastest rate possible will evolve toward a state that is increasingly complex, organized, alive, and intelligent*. Why? Because the computation that life must perform in order to persist indefinitely dissipates free energy more efficiently than any other natural process. Wissner-Gross calls the driver of this organizing process a "causal entropic force," and it ties together evolution and the second law.

The gist of the idea can be summed up easily enough. By becoming more diverse and complex through evolution, life maximizes the number of niches it can exploit, which optimizes its ability to extract energy and evade entropic decay. By seeking to preserve its future freedom, adaptive complexity grows more "empowered," which means it can respond to a greater diversity of challenges. As a result, life grows increasingly proficient at manipulating matter and energy in ways that will ultimately allow it to become a cosmic force of unparalleled causal power. The new cosmic narrative that is emerging from these insights depicts a universe that is growing increasingly intelligent through life's expansion, with no signs of slowing down.

Many believe that this process is unsustainable; that for it to go on indefinitely would go against the almighty second law, violating the heat death assumption, but we will see it is just that—an assumption. The heat death narrative, also known as the *big freeze*, is being quietly challenged by an increasing number of leading physicists, as physicist-turned-origin-of-life researcher Paul Davies explained: "We now see how it is possible for the universe to increase both organization and entropy at the same time. The optimistic and pessimistic arrows of time can co-exist: the universe can display creative unidirectional progress even in the face of the second law."[1]

But this statement is not quite strong enough. It would seem that the open-ended growth of organized complexity can *only* occur spontaneously

in a universe that otherwise becomes increasingly disordered. This is one of the fundamental insights of the integrated evolutionary synthesis, and such a truth could not have been discovered without thinking about evolution from a thermodynamic and informational perspective, as Schrödinger suggested. In this chapter, it will become clear why the tendency toward disorder is the selection pressure for self-organizing systems. It might sound complex on the surface, but again the story is surprisingly simple: the second law sculpts self-organizing systems to be increasingly stable by selecting the most synergistic structural states.

## One Cohesive Cosmic Story

The dogmatic reductionist, who much like the religious ideologue fears change and is invested in preserving the status quo, can be comforted by the fact that this radical extension of evolutionary theory is actually not that new and not that radical. Past attempts to achieve the same goal have generated a robust body of empirical and theoretical work over many decades, and it is this body of peer-reviewed research that informs the present approach. In fact, we can see the progressive attempts at a unified systems theory—cybernetics, general systems theory, evolutionary epistemology, universal Darwinism, and complexity science—as a striking demonstration of the evolutionary process they are all trying to explain. Paradigms emerge and evolve just like organisms and cultures. In this way, the integrated evolutionary synthesis, the unifying theory of reality developed throughout this book, predicts its own emergence. In a mechanical sense, cosmic evolution was destined to produce a theory of cosmic evolution. Although it goes back at least to Darwin's famous contemporary Herbert Spencer, an evolution-based worldview became a serious philosophical position with Karl Popper.

Popper saw that the scientific method was really just a new spin on what nature has been doing for billions of years—hypothesis testing is an implementation of an algorithm first discovered by evolution, though later it would be co-opted by the intelligent beings that the mechanism produced. It has different names depending on the field of application: conjecture and refutation (science), trial and error (learning), variation and selection

(adaptation), and generate and test (AI), but they all refer to the same general algorithm, which when applied iteratively performs Bayesian inference.

Identifying the algorithm that is driving cosmic evolution and knowledge creation does not mean we are done. The point of a proper theory of everything should not be to say, "We've found a master mechanism—the mystery of existence has been solved!" Remember our aim is not to reduce, but to illuminate, and variation and selection is simply the problem-solving procedure through which adaptive complexity finds solutions to survival problems and writes them into memory. While it is the general algorithm that must be at work anytime there is any functional design in nature, it is still just a procedure. We care about it only because it is the mechanism that produces knowledge. But why do we care about knowledge?

Knowledge is interesting to us because it is what allows physical systems to be able to experience the world and to generate patterns of activity that have meaning and feeling, thereby transforming nonliving parts of the universe into pieces with purpose. This is not just a metaphor; it is literally true that when inanimate matter acquires the ability to sense and perceive reality by arranging itself into adaptive complexity, matter becomes "alive," animated by the information processing that drives the behavior we call agency.

But if we focus solely on the individual organisms, or the complex adaptive system we know best—that is, ourselves—without thinking about the larger cosmic process at work, then we will be blind to the big picture. The continual growth of knowledge is infinitely intriguing because it appears to inevitably lead to a universe that organizes itself into an increasingly intelligent computational agent, with no apparent end in sight. Is the ultimate result of this process a cosmic mind that grows more powerful with time? The cosmos has created a puzzle for self-aware systems to solve. The riddle, the one we are unraveling right now, the question of how long life can last in a universe that obeys the second law, is both the problem and the solution. That is to say, it is our awareness of impending doom that motivates us to evade such an outcome. In this way, problems push progress forward.

# Poetic Meta-Naturalism

Our unifying theory of reality, the integrated evolutionary synthesis, inspires a new scientific worldview and existential philosophy that we may call *poetic meta-naturalism*. It does not have a flowery name for the sake of being poetic; there are technical reasons for this exact choice of words that will become increasingly clear as this book moves toward its conclusion.

Fans of the theoretical physicist and popular author Sean Carroll, a self-described "poetic naturalist," may be familiar with his philosophical position *poetic naturalism*, which was beautifully articulated in his ambitious 2016 book *The Big Picture*. Although there are similarities between poetic naturalism and poetic meta-naturalism, in some ways, they could not be more different. Like the stance of Brian Greene, Carroll's position is poetic but still overly reductionistic. It is clear that he does not believe the universe has any intrinsic purpose or that life has any cosmic significance:

> A poetic naturalist will deny that notions like "right and wrong," "purpose and duty," or "beauty and ugliness" are part of the fundamental architecture of the world. The world is just the world, unfolding according to the patterns of nature, free of any judgmental attributes. But these moral and ethical and aesthetic vocabularies can be perfectly useful ways of talking about the world.

Poetic meta-naturalism, on the other hand, argues that nature has an intrinsic purpose—to wake up and experience the fruits of its own creation. To do so, it must produce life, knowledge, and consciousness. The universe is inherently creative, and the emergent phenomena it generates are its poetic expressions. It is not the particles that are significant in the grand scheme of cosmic evolution, but the patterns that emerge from particle interactions. Some of these patterns, the ones with the most causal power over nature, will experience what it is like to be a pattern. You are such a pattern. Patterns are just as "real" as particles, and that includes actual poems. A poem that inspires an action is just as real as a rock.

According to poetic meta-naturalism, we can better understand the true nature of nature by thinking about the evolutionary process in terms of *loops and levels*. Without these concepts, it would be impossible to understand key properties of biological systems, like control and consciousness. We must thank the pioneers of cybernetics for making them a proper part of science. Armed with the logic of loops, we can begin to develop a science of the self.

## We Are Loops

In his essay for the digital magazine Aeon, titled "Consciousness Is Not a Thing, but a Process of Inference," neuroscientist Karl Friston explains our loopy nature and its relation to the second law:

> What's remarkable about this sort of repetitive, self-organising behaviour is that it's contrary to how the Universe usually behaves. Everything should actually get *more* random, dispersed and chaotic as time marches on. That's the second law of thermodynamics—everything tends towards chaos, and entropy generally increases. So what's going on?
>
> Complex systems are self-organising because they possess *attractors*. These are cycles of mutually reinforcing states that allow processes to achieve a point of stability, not by losing energy until they stop, but through what's known as *dynamic equilibrium*.

Notice that in this context, dynamic equilibrium is the *opposite* of thermodynamic equilibrium, a kind of non-equilibrium equilibrium! This sounds paradoxical on the surface, but it is just a quirk of language. While thermodynamic equilibrium is a balanced state of total disorder and energy dispersal within a system, dynamic equilibrium refers to a balanced state between an open system and an environment that is feeding it a steady supply of free energy. Both states are types of attractors, but one represents death while the other is associated with life. Dynamic equilibrium, a far-from-equilibrium steady state, was described mathematically by Ashby in 1947, when he came up with his *principle of self-organization*, which says

that "any deterministic dynamic system automatically evolves towards a state of equilibrium that can be described in terms of an attractor in a basin of surrounding states."[2] Once an adaptive system falls into an attractor, it will remain there as long as it continues to extract the energy it needs to sustain its ordered state, and as long as external conditions do not change too drastically.

Drawing inspiration from Ashby, Karl Friston explained that attractors are at the very heart of what it means to be alive:

> An intuitive example is homeostasis. If you're startled by a predator, your heartbeat and breathing will speed up, but you'll automatically do something to restore your cardiovascular system to a calmer state (following the so-called "fight or flight" response). Any time there's a deviation from the attractor, this triggers flows of thoughts, feelings and movements that eventually take you back to your cycle of attracting, familiar states. In humans, all the excitations of our body and brain can be described as moving towards our attractors, that is, towards our most probable states.

Continuing, Friston characterizes the living state as a loop, citing the cognitive philosopher and Pulitzer Prize–winning author Douglas Hofstadter, who we will get to know better in Part Three:

> On this view, humans are little more than "strange loops," as the philosopher Douglas Hofstadter puts it. We all flow through an enormous, high-dimensional state-space of manifold possibilities, but are forced by our attractors to move around in confined circles. We are like an autumn leaf; tracing out a never-ending trajectory in the turbulent eddies of a stream, thinking our little track is the whole world . . . This description of ourselves as playful loops might sound teleologically barren—but it has profound implications for the nature of any complex system with a set of attracting states, such as you or me.

Not teleologically barren at all! As complex adaptive systems, there is no doubt that we are loops, and in more ways than one, as this story has

shown. As dissipative systems that constantly cycle energy through metabolism, we are a kind of thermodynamic loop, and because we update our beliefs about the world in a Bayesian manner, we are also an informational loop. Additionally, we are cybernetic machines that use positive feedback loops for self-regeneration and self-amplification, and negative feedback loops to return to homeostasis and to shield the system from heavy perturbations that threaten its physical integrity. Loops are at the heart of who we are. Even the way we live, the patterns of behavior, the daily routines—breakfast, lunch, dinner, sleep, sex—our habits are attractors, for better or for worse. We are recursive creatures at our core.

But the loopy nature of nature shouldn't at all make us feel like we are eternally going in circles, because the loop is a Bayesian model update and the method of knowledge creation. Be it through science, adaptive learning, or biological evolution, life learns from its mistakes and records its solutions in memory—whether genomes, brains, cultures, academic journals, or web pages—so that they are not forgotten. From the continual feedback that we get from each of our failures and successes, the learning loop pushes the process in the direction of progress.

At this point in the story, it is time to introduce an important phenomenon that comes along with loops known as *self-reference*.

## Self-Reference

A helpful rule of thumb is that any time there is the word "self" in a scientific term, there's a good possibility that self-reference is lurking somewhere in the concept. Self-organization (emergence), self-maintenance (autopoieses), self-amplification (growth), self-replication (reproduction), and even self-awareness (conscious of being conscious) all involve a recursive process that utilizes loops in some way or another, allowing the system to reference itself.

All organisms, including humans, harness feedback loops to exercise causal power and control. The same kind of nonlinear, exponentially growing feedback loops that underlie the famous butterfly effect are channeled by life to do things that inanimate matter could never be capable of, like moving against the force of gravity the way we do when we run

uphill or jump up toward the sky. If you saw a nonliving object, like a stone or trash can, move in the same way, you'd assume that someone was playing a trick on you. Adaptive systems are capable of doing what would appear to be "causal magic" from the classic reductionist perspective. Of course, that magic is just the causal power of information, but that doesn't make it any less magical, does it?

Perhaps the most interesting thing about our unifying theory of reality is that when you realize it is a theory of adaptive systems, the theory suddenly becomes about you, pulling you into the loop! You are a loop inside a loop.

"Emergence has in an orderly way moved from protons to philosophers," wrote the Yale biochemist Harold Morowitz, founding editor of the journal *Complexity*. "At this level there is a kind of closing of the loop, because philosophers think about Big Bangs, protons, and all other hierarchies connected by emergences. The emerging world turns inward and thinks about itself."[3]

Harvard astrobiologist Eric Chaisson, author of *Cosmic Evolution: The Rise of Complexity in Nature* and longtime proponent of the idea that the cosmos itself is self-organizing, says, "Cosmic evolution has brought us forth, and now, having done so, enables us to study that very same scenario of cosmic evolution."

Along similar lines, Paul Davies wrote: "The laws of the universe have engineered their own comprehension."[4] It requires some reflection to grasp, but it's true, and they achieve it through beings like you.

As mentioned, a theory of cosmic evolution would also be a theory about the evolution of paradigms, meaning the theory would refer to itself. This instance of self-reference may sound strange on the surface, but Hofstadter has said many times, "a theory can talk about itself" in his strange and brilliant bestsellers *Gödel, Escher, Bach: An Eternal Golden Braid* and the heartfelt *I Am a Strange Loop*.

Our unifying theory of reality, with its basis in Bayesian epistemology, predicts that its model of the world will have errors, and it gives instructions for how it is supposed to self-update when new evidence comes in (using Bayes' theorem). Since the theory is self-correcting by nature, you could call it "self-aware." It is also by definition an *incomplete* theory, as it acknowledges that new phenomena with novel properties will inevitably

emerge in the future, requiring that the framework be expanded. Emergent phenomena will always exhibit new kinds of dynamic behavior that will require new mathematical laws to capture—like we've seen with the invention of causal calculus by the Israeli American computer scientist Judea Pearl, which has been used to formalize the emergence of causal power in living systems.

By declaring that reality is "open," meaning that novel patterns and experiences will be continuously created, poetic meta-naturalism shows us that a model's incompleteness should not be mistaken for a weakness. A theory predicting that reality is incomplete is in some sense complete, because it can predict what it can't predict perfectly. We can expect emergences of increasing grandeur, but those emergences will always contain surprises.

As we get sucked into the loopy story, we realize that this universal theory of cybernetic systems suggests that the universe itself is a self-organizing cybernetic system, rather than just any old arbitrary physical system, as Laplace and Monod characterized it. The emergence of a unifying theory signifies that we are at a critical point in human history, because for the very first time we are becoming aware of our potential starring role in the great unfolding cosmic drama that awaits our civilization, should we choose to correct our errors before we become the errors that get corrected. The universe is self-organizing, cosmic complexity is increasing, and intelligent life is the primary driver of the process. How does the universe "wake up"? Through loops, we now know, but loops alone won't suffice for mind. Sentient beings can only be understood as a product of self-reference and a multilevel informational loop that Douglas Hofstadter calls a "strange loop," which is not just a loop—it is specifically a loop that crosses *levels*. Levels are just as important as loops to the new cosmic narrative.

## We Have Levels

As complex adaptive systems, we have levels. In other words, all adaptive systems are nested systems: systems made of many interacting subsystems that are themselves made of many interacting subsystems. Cells are made of

biomolecules, multicellular organisms are made of cells, and eusocial societies, like an ant colony or a human civilization, are made of multicellular organisms. While we are wired to think of ourselves only as individuals, we, too, are communities of agents, as any cellular biologist would tell you (the human body is an integrated network of about thirty-seven trillion cells).

This nested, many-component design is called a *hierarchical modular architecture*: "modular" meaning the system is made of many interacting units, and "hierarchical" meaning that the components are organized into distinct layers or levels. It's units made of units all the way down. This modular multilevel architecture is recognized as being the most robust design an adaptive system can have—for a number of reasons. For one, since the computational load is distributed among many components and across many levels, there is a kind of specialization at each layer. But beyond more computational power, a hierarchical modular design can protect the system from being killed or destroyed as easily.

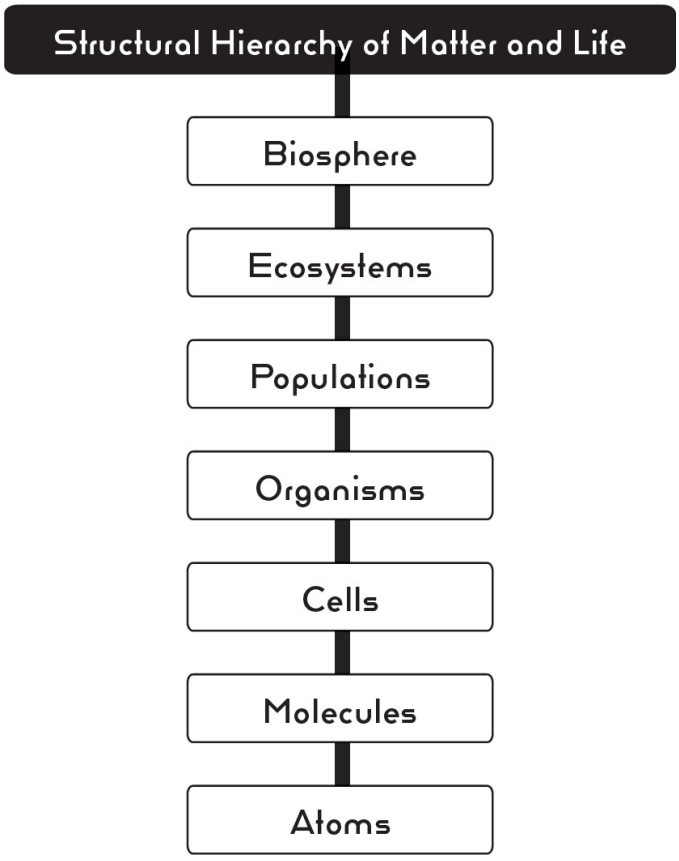
Consider a society of people, which is also an adaptive system, and a cybernetic system with a hierarchical modular architecture. The functional similarities between organisms and societies have inspired a number of well-respected scientists, like the esteemed biologists E.O. Wilson and Lynn Margulis, to describe the integrated collective as a "social organism." Even the great evolutionary biologist Ernst Mayr, who was an outspoken critic of the idea of a cosmic teleology, was adamant that "the social group is also a target of selection."[5] The resilience of social organisms makes it very easy to see the benefits of a nested, multilevel system design.

For example, if the higher level system falls apart—due to famine, natural disaster, war, or any other catastrophic disruption—the subsystems that make up the larger system can simply reassemble. In the case of failed civilizations, human or insect, the autonomous agents will eventually explore new collective configurations until a more stable and synergistic pattern of arrangement is discovered and retained. A hierarchical modular system is an attractor built out of attractors.

This nested architecture explains the biosphere's robustness, and why it can continue to carry out its thermodynamic function even in the face of natural disasters and mass-extinction events. The first scientist to

thoroughly explore the importance of hierarchical modular design was Herbert Simon, an American economist, political scientist, and cognitive psychologist who won the Nobel Prize in economics in 1978. His work influenced numerous fields, especially the young field of artificial intelligence. Simon's ideas about the importance of hierarchical modular design helped shape the field known as *control theory*, also called *optimal control theory*, which is now part of the core curriculum for engineers trying to develop robust systems of all kinds. The concept has also become a central part of machine learning, especially in the design of neural networks.

The figure on the next page represents the hierarchy of matter, which you can see continues as the hierarchy of life.




From this depiction, we can see that once life emerges, it becomes the driver of cosmic evolution, and we also clearly see that the story of cosmic evolution is a tale of hierarchical emergence through recursive self-

organization. We know that self-organization is how the levels in the hierarchy get built, but what we really want to know is if this process will continue into the future. Is evolutionary progress as inevitable as life's emergence? It certainly seems so, and we will describe the open-ended process with a law inspired by Ashby's principle of self-organization, called the *principle of recursive self-organization*. This principle will explain how new levels get added to the ever-extending hierarchy of matter through evolution. These levels will take us from matter to mind and beyond.

The saga of cosmic evolution is a recursive process of spontaneous self-organization that occurs in stages, and each stage is a "major evolutionary transition" that represents the emergence of a new level of structural and functional complexity in the universe. New levels in the hierarchy represent a higher level of control and causal power (once fully developed), and with each major evolutionary transition, some entirely novel property emerges that was not present in the systems below when they existed in isolation. These transitions were first described in detail by John Maynard Smith and Eörs Szathmáry in the 1995 book *The Major Transitions in Evolution*.

Typically, a major evolutionary transition involves the creation of a new kind of knowledge repository, which is a memory system that stores adaptive information. The first knowledge repository was a single cell, then later came multicellular organisms, then brains, then cultures, then computers, and now a "global brain" seems to be emerging, thanks to the internet. Some of these major evolutionary transitions correspond to what philosophers call *strong emergence*.

The more conservative brand of emergence, known as *weak emergence*, is all throughout nature wherever there is pattern, and it is the kind of emergence we see when dissipative structures form, like tornadoes and whirlpools, which do not have agency. Strong emergence is a special kind of emergence that creates the existence of something ontologically new in nature. The origin of life counts as a strong emergence because it is when information gains causal power in the universe, and the origin of mind also makes the cut because it represents the emergence of subjective experience, which is something altogether new.

Each emergence in the cosmic self-organization process is brought about by a phase transition that moves life ever further away from a state of thermodynamic equilibrium and total disorder. Through a nested series of

such phase transitions—where functional things come together to make larger functional things, which come together to make even larger ones, and so on—adaptive complexity becomes better equipped to dominate the cosmos. That is, life acquires an organization that is increasingly hierarchical and integrated, and therefore more resilient and computationally powerful. While it is okay to describe life as dominating the universe by gaining dominion over it through technological progression, it is just as true that the universe itself is coming to life through this process. It is the same phenomenon described from different perspectives corresponding to different levels of description. Remember Sagan's slogan, "We are a way for the cosmos to know itself."

Now we will illuminate the nature of this process by describing the precise mechanisms that make major evolutionary transitions inevitable. Like the first major transition in cosmic evolution, which featured the emergence of a living organism from the coordinated collective activity of many interacting molecules, the major evolutionary transition that is approaching—brought about by the coordinated collective activity of trillions of different kinds of interacting species—represents the developmental maturation of the biosphere into a true "superorganism." This planetary-scale complex adaptive system has a brain, a body, and potentially, the ability to reproduce.

## Evolving Beyond Earth

Cosmic evolution didn't begin on Earth of course; it started when the universe itself began. We know that the formation of the non-equilibrium structures we call stars and planets happened inevitably as a result of the fundamental laws of physics, and we are totally comfortable with the inexorableness of that part of the story. But the idea that life is the same kind of regularity is not currently the position of mainstream physics, even though it is certainly the favored view among complexity scientists, NASA employees, and a growing number of bored physicists looking to explain life's origin, a new holy grail for those schooled in statistical thermodynamics. From Nobel Prize–winning biologists like Christian de Duve to our most trusted science communicators like Carl Sagan, the idea

that life's emergence was "written into the fabric of the universe," as de Duve put it, is no longer shocking.

While the idea that life and agency emerge as an inevitable result of natural law is something we should find spiritual satisfaction in, it leaves us wondering where the inevitability ends. Even though it implies life in the cosmos is common, readily emerging on those billions or trillions of planets with sufficiently Earthlike conditions, it would be quite disappointing if this only led to life that could never leave its planet of origin.

If biology is to expand beyond its planet of origin, then it must evolve intelligence, and intelligence is a property that emerges in sufficiently complex information-processing systems. So, if this cosmic trajectory is in fact obligatory given the laws, then there must be mechanisms that systematically drive the biosphere to evolve toward states of higher complexity and information integration. In the previous chapter, we learned a lot about the knowledge-creation process, but we didn't talk much about how the hardware that encodes the growing mountain of accumulated knowledge gets upgraded. Major evolutionary transitions are revolutions in machinery for information processing and storage. The need for ever-larger knowledge repositories (memory systems) reveals that the story of cosmic evolution is not just a process of knowledge creation but a process of *recursive self-organization*.

To understand the principle of recursive self-organization, we have to think about how the two types of evolution we've identified, which correspond to two types of learning, work together to build cybernetic systems of increasing computational power. We are now pretty familiar with phylogenetic learning and the basics of ontogenetic learning, or self-organization. As it was used traditionally, the term "ontogenetic learning" exclusively meant individual learning, but as science has grown more sophisticated, it has become clear that all the things we call individuals are actually self-organized collectives of smaller units that work cooperatively. Therefore, we will use ontogenetic learning to describe any example of self-organization, as it is the "self" in each instance that makes collective learning individual learning on a higher level!

While the modern synthesis emphasizes how evolution occurs as a result of competition between organisms fighting for survival—the source of phylogenetic learning—it says very little about the coevolution of

species and the interdependence and interconnectedness we see in the biosphere, which has a resilience and stability that must be explained. Phylogenetic learning is the method of Bayesian updating that optimizes the genome of an evolving population of organisms, but self-organization, the lesser known of the two evolutionary processes, really is the star of the show, because it is so widespread, and because it is the primary mechanism driving continual emergence.

Despite its obvious relevance, self-organization has been largely ignored by mainstream biologists, for both practical and cultural reasons. Not only do biologists tend to avoid concepts that require exotic mathematics and physics to characterize, they also generally prefer to keep a healthy distance from anything that sounds too teleological for comfort, as they still associate teleology or goal directedness with some mystical or "nonphysical" force that could let God or magic slip into the scientific enterprise through the back door. Well, those careful guardians of evolutionary theory—a sort of "Night's Watch" for science—can relax, celebrate even, because although self-organization is a process that mystifyingly brings about a "goal state," that goal is an attractor with a thermodynamic function and a precise mathematical description. Not only that, naturalists will be pleased to know that this attractor is actually established and maintained by a bona fide Darwinian dynamic, as it evolves through variation and selection.

## Blind Variation and Selective Retention

More specifically, evolutionary epistemology shows us that self-organization occurs through the variation-and-selection mechanism known as blind variation and selective retention. We can think of this version of the Darwinian dynamic as an expression of Dawkins's statement that "Darwin's 'survival of the fittest' is really a special case of a more general law of *survival of the stable*." When a system of interacting parts is exploring various configurations, the stable ones get retained or selected while the unstable ones are forgotten.

The first time the mechanism of blind variation and selective retention makes its presence obvious is with the origin of life, in the form of what

Jeremy England has called "dissipative adaptation." We learned all about this process in Chapter Three, but from this it was not made fully clear that dissipative adaptation through self-organization was a *multiply realizable phenomenon*. This means that because functionally equivalent complex adaptive systems can exist on different types of hardware, or "substrates," the same type of phenomenon or process can be "realized" at multiple scales in nature.

*Multiple realizability* is also known as *substrate independence*. Because of substrate independence, studying a complex adaptive system at one level can lead to insights that help us better understand similar adaptive systems at other levels. In 2017, physicist Max Tegmark said the concept that he thought society should know more about was substrate independence, and he then proceeded to explain how a scientist can understand a wave independently from the medium it travels in. Waves behave similarly no matter the medium because the phenomenon of interest is to be found in the dynamics, not the material. Complex adaptive systems are cybernetic systems that can exist on an infinite number of scales, in theory.

## The Meaning Is in the Metaphor

Poetic meta-naturalism stresses the importance of using metaphor in explaining natural phenomena, not for the sake of being artistic, or relatable, but because it becomes a necessary explanatory tool in a reality with levels. When we are trying to understand the nature of a multiply realizable phenomenon, the "true" description of that phenomenon is not to be found at any single level—the deepest understanding would come from recognizing the common abstraction that can be extracted from the different levels. For example, if information theorist Erik Hoel is right in his belief that "agenthood seems to be a multiply realizable property,"[6] then agency should be defined as the property that is shared by agents at all levels of scale.

In mathematics, there is the concept of *self-similarity*, and a self-similar object is one that is similar in structure to one or more (or all) of its parts. Self-similarity is associated with fractal patterns, which emerge from

recursively applied rules of design, and they exist all throughout nature, often in the form of the famous Fibonacci sequence, which we see in the seed arrangement of sunflowers, the spirals of galaxies, and the social organization of honeybee colonies. Systems that show self-similarity, like cybernetic systems with a hierarchical modular architecture, can in theory be understood in new ways through new kinds of high-level principles, like what mathematicians call a *reflection principle*.

A reflection principle proves that the levels in a nested system can inform each other by revealing dynamics that might be hard to observe, for whatever reason, at some other level. To use an example we will soon explore, the dynamics of ecosystems might tell us something about the dynamics of autocatalytic sets, and vice versa. One might also assume that learning about the dynamics of the complex adaptive system at the highest level of the growing hierarchy of life might tell us something about the levels that *have not yet emerged* but are to be expected. For example, studying how consciousness emerges in human brains could tell us something about how consciousness might emerge in the global brain that is being formed by humans connected by their wireless devices. In this way, poetic meta-naturalism has predictive power that no reductionist theory has. Rather than ignoring emergent phenomena, it uses them to predict the behavior of isomorphic patterns at higher scales.

The best way to illustrate the concept of multiple realizability is through example. In Chapter Six, we saw that by learning about science, we were also learning about evolution, and by learning about evolution, we were learning something new about life's origins. Now life's origins will teach us something we didn't know about life's future. Owing to the truth of multiple realizability, the dissipative adaptation model of abiogenesis is also the model for the evolutionary development of the biosphere.

## The Biosphere Is an Autocatalytic Set

Recall from Part One that a new understanding of how complex adaptive systems originate, based on the new field of physics known as non-equilibrium thermodynamics, tells us that life arises naturally and predictably given certain geochemical conditions that are not uncommon in

the universe. Through dissipation-driven adaptation, a system of organic molecules pushed by a steady flow of energy will spontaneously rearrange its configuration into an organization that creates a self-renewing, self-amplifying chemical reaction network called an *autocatalytic set*. This is an example of emergence, and it occurs when the system undergoes a phase transition to a state of higher order, complexity, and stability.

In his 2006 book *Programming the Universe*, physicist Seth Lloyd characterizes the autocatalytic set as being like a factory powered by the energy that flows through the primitive metabolic system: "Autocatalytic sets of reactions are powerful systems. In addition to computing, they can produce a wide variety of chemical outputs. In effect, an autocatalytic set of reactions is like a tiny, computer-controlled factory for producing chemicals. Some of these chemicals are the constituents of life."

Chemical autocatalytic sets are the kind we usually read about, but an ecosystem does collective autocatalysis too. Stuart Kauffman, who has no doubt done the most to make the autocatalytic-set theory of life's origin the leading theory, emphasizes that the creation of new species and new niches is no different in any mechanistic sense from the kind of collective autocatalysis we see in relatively simple chemical networks. In the 2018 paper, "Niche Emergence as an Autocatalytic Process in the Evolution of Ecosystems," published in the *Journal of Theoretical Biology*, Kauffman and colleagues describe why ecosystems are "autocatalytically closed self-sustaining reaction (or interaction) networks" that reliably drive up biological diversity and complexity as they self-amplify and evolve.

Quoting the cyberneticist Francis Heylighen quoting E.O. Wilson—call it a "meta-quote"—we see that ecosystem-level autocatalysis is continual and seemingly open-ended. In his 1999 paper "The Growth of Structural and Functional Complexity During Evolution," Heylighen said, "As is well explained by E.O. Wilson (1992), not only do ecosystems contain typically lots of niches that will eventually be filled by new species, but there is a self-reinforcing tendency to create new niches." If this sounds familiar, it's because we learned about it in the previous chapter.

When you think about it, almost all integrated networks of adaptive systems function as autocatalytic sets, including the social organisms we call societies or civilizations. Such networks are common because they emerge spontaneously when many organisms repeatedly interact with one

another and discover synergistic collective configurations, as nature often pressures them to do.

Why do interacting agents consistently link up to form stable wholes? For the exact same reason that molecules with the right chemical diversity will form stable autocatalytic reaction sets when pushed by a flow of energy: Working collectively allows the whole system to extract more free energy with less work. This is the essence of the principle of recursive self-organization. Nature promotes cooperation, collaboration, and synergy because it is thermodynamically beneficial for all parties, and for that reason, synergistic collective configurations will eventually be discovered by any many-component system that is exploring various states or configurations through the blind-variation-and-selective-retention mechanism. Organisms only compete until they finally figure out that working together makes everyone's task easier, and that goes for humans too.

As long as a biosphere is creating a growing variety of complex adaptive systems, some of those systems will interact to produce higher-level complex adaptive systems, which will come together to form even larger complexes, and this process continues at higher and more computationally sophisticated levels as time marches on. Where the chain of emergence ends, or if it ends, no one knows, but we will have some fun speculating about that in the final chapter.

Familiar examples of so-called superorganisms include ant colonies, termite mounds, human societies, and the ecosystems that make up the biosphere. All of these distributed networks process information and fight to maintain their organization in ways that are surprisingly similar to the organisms they are made of. While the name "superorganism" is provocative, it is important to recognize that a society or civilization is just the next emergent level of structure and function in the hierarchy of life that is being built up by recursive self-organization. Every organism is a superorganism from the perspective of the level below, and every superorganism is just an organism from the level above it. For example, if a single-celled creature like a bacterium could reason, it would consider multicellular organisms like humans to be superorganisms.

Superorganisms have also been called meta-organisms, metasystems, and supersystems, but the general idea underlying all these supercool names

is the same: They are complex adaptive systems made of many complex adaptive systems. The parts come together to form greater wholes through phase transitions, a process that has been described in mathematical detail by Eric Smith and Harold Morowitz in their 2016 book *The Origin and Nature of Life on Earth*, which proposes the phase transition theory of life. This theory is similar to Valentin Turchin's metasystem transition theory. Turchin was a twentieth-century Russian computer scientist who described in the 1970s how the process of recursive emergence would inexorably create increasingly complex cybernetic systems. The theory was outlined in his underappreciated cybernetic epic, *The Phenomenon of Science*, a title that was an obvious tribute to Pierre Teilhard de Chardin's *The Phenomenon of Man*. In this picture of cosmic evolution, the universe is assembling itself into a hierarchical (multilevel) cybernetic machine with godlike powers.

## Synergistic Collective Configurations

Recall that phase transitions occur when the interactions between component parts, whether molecules or agents, become coordinated enough to create cooperative effects, also known as mutually reinforcing effects, that push the system to cross some critical threshold. How does this magical metasystem transition happen exactly? Well, it starts with a flow of energy that drives the many-component system toward an attractor in its configuration space. When one component interacts with its neighbor, it produces an effect in that neighbor, which gets propagated to its surrounding neighbors, which propagate the effect of the initial cause further, creating a loopy chain reaction. Since the whole network is dynamically interacting, the causal chains produced by global propagation eventually loop back to their points of origin. In effect, the process creates a repeating series of causal loops that entrain the parts of the newly formed whole, harmonizing collective patterns of motion to allow for more efficient dissipation of an energy gradient.

Per the constructal law introduced in the prior chapter, this enables even more energy to flow through the system, organizing it faster, such that a positive feedback loop amplifies the size and complexity of the network.

Out of molecular or behavioral chaos, dynamical order spontaneously emerges as new functional structure crystalizes from a quite enchanting entropic force. The individual components are now part of a higher-level system, with new global properties that constrain the behavior of the elements that make up the system.

While some of the individual freedom of each molecule or agent is lost as it becomes part of a collective, the synergy gained from joining the group affords the individual new freedoms undreamed of before. Individuals who are part of a society with rules may be less free than a person living in the wild, but they have access to the fruits of civilization, like technology that can extend human experience. Without culture, humans cannot know what it's like to fly in a plane or listen to a symphony. Without technology, life cannot get off the planet or evade the death of its star.

Although spontaneous self-organization is a process that is intrinsically teleological, or goal-oriented, as the dynamical system inevitably enters an attractor that it must actively maintain, it is equally Darwinian. As the energy flowing through the system drives it to continually explore new configurations, the states that best capture and dissipate energy are the ones that the system gets repeatedly stuck in. This is a form of natural selection.

Remember, fitness is really about stability, so the most stable states are selected by the second law—or the tendency toward disorder. As the evolving system moves toward its attractor and becomes increasingly stable, adaptive information is being accumulated in its pattern of configuration. As a result of self-organization, or ontogenetic learning through blind variation and selective retention, the techniques for evading equilibrium get encoded in memory, which is stored in the dynamical organization of the adapting system. Remember Susanne Still's statement from the paper quoted near the end of Chapter Six: "A system responding to a stochastic driving signal can be interpreted as computing, by means of its dynamics, an implicit model of the environmental variables. The system's state retains information about past environmental fluctuations, and a fraction of this information is predictive of future ones."[7]

This applies to self-organized systems at all levels, whether they're cells, brains, or cities. Karl Friston and Maxwell Ramstead of McGill University have published papers with colleagues that work out the math

for how the free energy principle is to be applied at all these levels. Here we learn an important lesson: When interacting agents self-organize into a synergistic collective configuration, that process is always a form of dissipative adaptation, because the emergent system can suck more energy from the environment. But dissipative adaptation is also a process of Bayesian inference, and non-equilibrium attractors are sets of states that minimize the system's free energy and prediction error.

What does it all mean? It would seem to suggest that the biosphere is following a similar evolutionary trajectory as the autocatalytic set that led to life, just on a larger scale. At first it is only a basic cybernetic system, but billions of years later it becomes an organism with a global brain and the ability to self-replicate. How does the biosphere reproduce? When we terraform other planets, like Mars, that is the biosphere reproducing! But we are not at that part of the story quite yet.

## All Species Matter

As we look back at the great teleology debate, we see that Stephen Jay Gould and his secular but spiritual critics like Robert Wright can be correct at the same time—evolution through natural selection is continually creating more complexity at the level of the biosphere precisely because it is continually creating more diversity at the species level. In fact, a biosphere fundamentally cannot grow more complex unless it is also growing more diverse, as the diversity of species is what creates a greater variety of specialized parts, and it is the variety or distinction of parts that allow for a division of labor among the collective whole. So, the growing ladder of niches of increasing complexity creates a distribution of interacting species that act as a variety of interacting parts, not unlike the variety of elements that are required to produce life.

More species in an ecosystem or biosphere is functionally analogous to more biomolecules in a cell or more cell types in a multicellular organism. This makes a lot of sense if you think about how other highly complex systems are structured. Neither a car nor a human would work well if each part were maximally complex, or if each part were evolving toward similar complexity. Remember, there is strength in diversity, and this goes for

systems of all types, whether they are biological, social, or technological. A car made of only engines would be a useless pile of junk, and a human made only of brains would be equally hopeless. The most functional systems have a range of parts that span a spectrum from simple to complex. As Darwinian evolution transforms life from a single cell into an ever-growing variety of interacting species and ecosystems, the biosphere develops from an embryonic stage into a mature superorganism with tissues and organs, as well as a nervous system that engenders it with a brain-like intelligence. We call this nervous system "the internet," and it includes the human minds it connects.

In a computational sense, humans are the "superior" species, but it isn't logical to think in these terms in any moral sense, because we are in no way independent from the biosphere we critically depend on, which includes every species still in existence. It would be like considering the neuron to be superior to the much simpler blood cell, or the brain to be a superior organ to the liver, whatever that would mean. Does it make sense to think about whether a car's engine is superior to the wheel? We need them all to have a properly functioning whole. So, Gould's fear of the idea of a ladder of progress is no longer a concern, because *the ladder is a web*.

The biosphere is a network of interacting adaptive systems that collectively create an interdependent, interconnected whole. There are no superior or inferior species, or superior or inferior humans. "I don't impute a value judgment to computational superiority," says chaos theorist Norman Packard,[8] who coined the phrase "the edge of chaos," which is where phase transitions to higher order occur. And neither should you.

An individual's value is not about physical ability or intelligence; it is about how one contributes to preserving the stability of the social organism we call society. By being unique, or having a unique perspective that you share with others through conversation, writing, or art, you are contributing to the diversity of ideas that collectively create the world model of the superorganism. For social organisms, we call that world model a *cultural worldview*—which can be a religion, national identity, political ideology, or cosmic philosophy. We can make use of metaphor by saying that *the worldview is the Markov blanket for the social organism*, because it is what binds the independent agents together to form a collective whole with

internal states that are distinct from the external states of the environment, even though there is no tangible physical barrier, like a cell membrane or an integumentary system. How do we know this is a justified analogy? Göbekli Tepe, that's how.

## Steps on the Cosmic Ladder

Göbekli Tepe is an archaeological site in Turkey with ancient ruins similar to Stonehenge, but much older, and with some important differences. Rather than just being big slabs of rock, the massive megaliths are carved into shapes that depict beings that are either creative renditions of humans and animals—or gods. It is thought that Göbekli Tepe was built in service to some worldview that symbolized humans coming together under some common goal or purpose, whether that purpose was service to a god or a more abstract sense of service to the cosmos that produced them. In the book *Sapiens: A Brief History of Humankind*, Yuval Noah Harari wrote:

> In 1995 archaeologists began to excavate a site in southeast Turkey called Göbekli Tepe. In the oldest stratum they discovered no signs of a settlement, houses or daily activities. They did, however, find monumental pillared structures decorated with spectacular engravings. Each stone pillar weighed up to seven tons and reached a height of five metres. In a nearby quarry they found a half-chiselled pillar weighing fifty tons. Altogether, they uncovered more than ten monumental structures, the largest of them nearly thirty metres across. Archaeologists are familiar with such monumental structures from sites around the world—the best-known example is Stonehenge in Britain. Yet as they studied Göbekli Tepe, they discovered an amazing fact. Stonehenge dates to 2500 BC, and was built by a developed agricultural society. The structures at Göbekli Tepe are dated to about 9500 BC, and all available evidence indicates that they were built by hunter-gatherers . . . Why would a foraging society build such structures? They had no obvious utilitarian purpose. They were neither mammoth slaughterhouses nor places to shelter from rain or hide from lions. That leaves us

> with the theory that they were built for some mysterious cultural purpose that archaeologists have a hard time deciphering. Whatever it was, the foragers thought it worth a huge amount of effort and time. The only way to build Göbekli Tepe was for thousands of foragers belonging to different bands and tribes to cooperate over an extended period of time. Only a sophisticated religious or ideological system could sustain such efforts.

The fact that Göbekli Tepe was created just a few thousand years before the first known civilization, the Sumerians of Mesopotamia, seems to suggest that the phase transition from small and loose hunter-gatherer communities to an actual civilization requires the emergence of some kind of worldview, because it is the worldview that aligns the interests of the selfishly acting individuals. What is the biological function of a cultural worldview? The worldview, like the world models that came before it, is a strategy for survival, but at this scale it is a strategy for *collective survival*, for the continued persistence of the superorganism.

So, here's the big picture. A biosphere will inevitably become increasingly complex and functional if there are evolutionary mechanisms at work that satisfy two basic requirements necessary for recursive emergence: the evolutionary process must generate an increasing variety of robust parts, and some of these parts must interact to form stable wholes. It seems that life is in luck, because the principles of Darwinian evolution and self-organization do just that, respectively. With a little help from the unintelligent designer we call natural selection, and the order-creating power of the phase transition, the biosphere magically churns out increasingly complex forms, creating a breeding ground for intelligence at both the individual and collective level.

Once we expand our definition of biological organization to include hierarchical architectures (multilevel complex adaptive systems) that exist above the level of the organism, such as social organisms, the trend toward higher complexity in the biosphere over time becomes impossible to deny. *Loops create levels, and these levels create steps on the cosmic ladder, with each new step representing an increase in memory storage capacity and computational power*. Acknowledging this trend toward higher complexity and control would be nothing less than a game changer, because it would

mean that the future would be semi-predictable. But to the scientist who has not updated the reductionist model of the universe to account for the causal power of information, recognizing the trend would amount to recognizing a property, namely teleology, which scientists see as being fundamentally opposed to naturalism, and that would be a big no-no. Who would have thought ideology and cultural factors would play such a decisive role in the specifics of evolutionary theory?

In his 2002 book *Life Evolving*, Christian de Duve tried to expose and correct this rigid ideology:

> A major objective of this book has been to expose the fallacy of [the] "gospel of contingency," which is being preached in the name of science. The alleged scientific premises of this doctrine, as I have tried to show, are incorrect. Not, as some would have it, because there is "something else" shaping the direction of evolution, but because the natural constraints within which chance operates are such that evolution in the direction of increasing complexity was virtually bound to take place, if given the opportunity. Chance does not exclude inevitability.

## The Gaia Hypothesis

A common and quite understandable objection to the teleological cosmic narrative is that it seems to ascribe conscious intent to biological entities and processes that aren't operating with any subjective awareness at all, like the biosphere as a whole and the evolutionary process itself. But this is not a problem for the modern notion of teleology put forth in this book, which is Darwinian and computational in nature. The levels of higher complexity that represent the major "goals" of cosmic evolution are not pursued consciously by the universe or by biospheres—at least not until intelligent, conscious agents emerge. They are simply attractors that spontaneously arise from the dynamics created by many interacting components or agents who are evolving as a collective unit.

These non-equilibrium attractors are states of stability that must be maintained by adaptive complexity if it is to advance to higher form, but the

individual agents themselves are typically not aware of their participation in a collective whole, nor is the collective whole typically aware of its role in constraining and regulating the behavior of its components. Therefore, we should view teleological goals as thermodynamic target states that use global feedback mechanisms to constrain and bind dynamically interacting components together into higher-order complex adaptive systems.

Although this certainly creates the appearance that there is something supernatural at work, the only magic that is needed is the magic of spontaneous self-organization and selection. As life blindly explores possibilities and selectively retains the solutions that move it forward, the living network collectively accumulates information about how to thrive in the environment over many generations and even "learns" new ways to evolve. In doing so, adaptive complexity unknowingly climbs a twisty teleological ladder of progress, becoming more robust with each setback and dead end it encounters along the way. What doesn't kill life not only makes it stronger but increasingly equipped to shape the path of cosmic evolution. In its perpetual struggle to stay far from thermodynamic equilibrium, adaptive complexity can't help but spread dynamical order in the universe.

Not only is life incredibly durable, the biosphere as a whole appears to function as a self-regulating system that is actively maintaining the conditions necessary for its own existence. This concept—that all biological life works together as a self-regulating system that is altering the planet for its own benefit—is known as the *Gaia hypothesis*. It was formulated in the 1970s by the British chemist James Lovelock, who was certainly thinking outside the box, perhaps with a little help from the psychedelic substances popular around that time. Gaia is the name of the primordial goddess of Greek mythology who is the personification of the earth, so you can think of her as "Mother Earth." At first, the theory was largely dismissed by biologists, mostly, if not exclusively, for ideological reasons relating to its mystical name. But the tides have turned, thanks to complexity science and computational neuroscience making cybernetics popular once again.

For what it's worth, Gaia was never meant to imply that the biosphere or planet was conscious or some godlike deity. Lovelock was simply saying that it is a self-maintaining cybernetic system, and an autopoietic entity. It was actually Lovelock's neighbor who suggested the name, and that

neighbor happened to be none other than William Golding, the Nobel Prize–winning author of the novel *Lord of the Flies*, so of course it is poetic and romanticized. While some might say a less loaded name would have avoided a lot of confusion and controversy, that may be true, but that controversy might be the reason anyone is talking about the idea right now. Marketing matters, and like it or not, "Gaia" grabs people's attention, probably for many of the same reasons "God" does, and that is okay.

Evolutionary biologist David Sloan Wilson, author of *This View of Life: Completing the Darwinian Revolution*, often compares the biosphere to a god and gets away with it because once you understand the computational intelligence of Gaia, then it becomes clear that the metaphor is more than appropriate; it is necessary for us to understand what kind of godlike powers a planetary-scale cybernetic organism can potentially have. Without analogy and metaphor, we have no frame of reference for understanding anything, because understanding comes from context.

A simple but powerful demonstration of the Gaia hypothesis comes from the idea that earlier life-forms on Earth made the planet habitable for the more complex forms of life that would emerge later, like the multicellular kind. Billions of years ago, the early atmosphere was utterly toxic for life as we know it. There was not much oxygen, and there were loads of poisonous gases like methane and carbon monoxide, so no animals that we are familiar with would have been able to exist. Fortunately for us, the photosynthetic organisms on Earth turned that noxious atmosphere into the one that allows beings like us to breathe, and Gaia keeps the specific gas composition stable for its complex carbon-based creatures. Not only that, it is well known that the biosphere acts as a sort of natural thermostat that actively keeps Earth's general temperature in a range that is friendly for life.

So now we can finally stop arguing about whether or not the biosphere, or a society, is an "organism," because we can give a formal mathematical and functional description of a general complex adaptive system, and we can see how all the examples of adaptive systems in nature differ and how they are alike. By recognizing the shared abstractions, we also learn what makes each new level in the hierarchy of life unique and special. Brains are kind of like organisms, and societies are kind of like brains and organisms, but each has different emergent properties and different dynamics. Once we map out the basic similarities and differences, the exact terminology and

metaphors we use become less important, because it becomes clear why and how certain language is being used.

## The Biosphere Becomes an Organism

At the present moment, the biosphere does not have all the properties that a traditional biological organism has, because it cannot reproduce, and in the minds of many evolutionary biologists, variation and selection can only happen through self-replication with genetic mutation. But we've learned that blind variation and selective retention is the Darwinian dynamic underlying self-organization, which is the process that powers the development of the organism from embryo to mature adult, and the development of the brain, an ontogenetic learning process described as *neural Darwinism* by the Nobel Prize–winning biologist-turned-neuroscientist Gerald Edelman. These mechanisms are all multiply realizable. So, our biosphere is not yet an organism, but it is an adaptive system and an autocatalytic set that is on its way to becoming one.

One could say that the biosphere technically becomes an organism the moment we populate another planet with adaptive systems that transform its surface into a biosphere that can support life, and this effort is already underway in the project that plans to terraform Mars. Once that happens, we may need to update the definition of an organism to something that is multiply realizable. What would define an "organism" would change as new iterations of the phenomenon come into existence.

Some might argue that successful societies self-replicate whenever a cultural worldview spreads to another part of the world and creates a functionally similar social organism, like when the idea of a democracy spreads. Physicist Geoffrey West's book *Scale: The Universal Laws of Life, Growth, and Death in Organisms, Cities, and Companies* is all about how adaptive systems at all the different levels, from cell to city, show common scaling laws and similar thermodynamic and computational relationships. So, the concept of a social organism is no longer just philosophy or poetry—it is a scientific term that expresses the truth of substrate independence.

We've talked a lot about the global superorganism, but does this superorganism have a mind? While the answer to that is uncertain, there is

no doubt that collectively, humans and their technologies are creating a planetary-scale nervous system—a superbrain for the superorganism.

## Meta-Evolution

We can think of the evolution of organization in the biosphere and society as a sort of meta-evolution: an evolution of metasystems made of many subsystems, but also an evolution of mechanisms of evolution. At higher levels, natural selection doesn't just act on genetic variation but on the variation of large-scale structure and processes. The human constructions that create or improve the social organism's stability are selected—such as good architectural designs, robust ideologies, and enlightening philosophies—while those that are not selected get weeded out by human civilization, through collective trial-and-error learning. This is another evolutionary factor that gives an overarching direction to transformative change. Through these processes, life becomes not only better adapted to niches, but better adapted to dealing with nature's challenges more broadly, and better suited to cope with the universe's conditions beyond the biosphere. It is through these evolutionary mechanisms that life can eventually become interplanetary, and someday, intergalactic. But for life to spread through the cosmos—for the biosphere to self-replicate and seed other planets—it seems that it must have a sufficiently developed brain. As the most complex species, the one that has mastered wireless communication, we are undoubtedly the neurons of the global brain.

Thanks to the internet, social media, and now blockchain ecosystems, intelligent agents both organic and synthetic are exchanging and integrating information through their digital and real-life communications, much like neuronal signaling in the nervous systems of organisms, but this nervous system spans the entire planet. In addition to at least four billion humans connected through internet access, there are also billions of AI systems and physical devices that are talking to each other and collecting and exchanging data across a network called the *Internet of Things*.

As a result of this technology-aided self-organizing process, a globally distributed superintelligence is emerging that will allow humanity to reach new heights. This collective intelligence is not separate from us, since we

are the nodes in the neural network that do the majority of the computing. But to fully realize this transformation, society must become aware of the larger consequences of our actions and the urgent need to coordinate them, so that we may foster global sustainability and promote progress for humanity as a whole. In doing so, the activity of the global brain will be making the biosphere even more robust and dynamic through new modes of self-regulation, increasing its causal power.

While most people are familiar with blockchain technology because of Bitcoin, we must recognize that this form of distributed computing is not just a vehicle for cryptocurrencies but an evolutionary force similar to the internet, though perhaps more disruptive and transformative. Blockchains and blockchain adaptations, like the directed acyclic graph, can connect humans in all sorts of new ways, spawning novel types of social, political, and economic networks.

These emergent systems increase the computational power that life wields, and their decentralized nature allows for improved systemic flow of resources and information throughout the superorganism. As the nodes in the human network become increasingly connected, the web of life becomes more resilient to the harshness of nature, and in many ways, better able to avoid existential catastrophe. If you are one who believes that humans are only capable of doing more harm than good to the organic life on this planet, think about the fact that human-level intelligence is the only thing that can save a biosphere with conscious creatures from a cosmic catastrophe, like a comet colliding with Earth.

The global brain can be seen as forming an "infosphere" on top of the organic biosphere—what Teilhard de Chardin called a *noosphere* ("noo" is Greek for mind), but can the global brain have an actual conscious mind? The gospel of multiple realizability would suggest that for a global mind to emerge from the global brain, its causal dynamics would have to duplicate the specific signaling dynamics of conscious brain states, as it is not just any old brain state that supports subjective experience, the property that defines a mind. Remember, consciousness can vanish while the brain keeps on ticking. So, it seems that before we think about whether the planetary-scale superorganism can be conscious, we must first understand how minds and mental models emerge from ordinary brains.

# PART THREE

# TRANSCENDENCE

*As chaos collapses into complexity*
*Chains of connected cells create cognition and consciousness*
*Conscious creatures coalesce into creative communities called cultures*
*And cultures create a collective consciousness that,*
*If cultivated correctly, will cross the cosmos*

*Since spontaneously self-organizing structures survive*
*So sturdy and solid, sentience shall surely spread*
*Suffusing space-time with self-awareness and subjective states*
*So stupendously sublime and surreal*

*Through our efforts to evade death and transcend our mortality*
*We get grander glimpses of the romance of reality*

# 10

# The Mind-Body Mystery

---

> Someday surely we will see the principle underlying existence as so simple, so beautiful, so compelling that we will all say to each other, "How could we all have been so stupid for so long?" I don't know whether it will be one year or a decade, but I think we can and will understand. That's the central thing I would like to stand for. We can and will understand.
>
> **—John Archibald Wheeler**

---

John Wheeler may be most famous in pop culture for coining the term "black hole," but I believe time will show that the physicist's most significant contribution to science and society was igniting the influential "it from bit" movement. This quiet revolution, catalyzed by Wheeler's 1990 paper "Information, Physics, Quantum: The Search for Links," puts forth the radical hypothesis that the universe is a giant computer.

What is this cosmic computer computing exactly? Its own future evolution. In this view, supported by Rolf Landauer's mathematical proof that "information is physical," the "things" of physical reality emerge from bits of information. Put playfully, *its come from bits*. These bits are encoded

in the states of particles, and when particles interact with one another and become correlated, nature is carrying out computations. As the cosmos computes, reality unfolds and simplicity gets transformed into complexity.

What this book aims to show is that the computational universe conceived of by Wheeler and popularized by Seth Lloyd is also the Darwinian universe described by universal Darwinism and evolutionary epistemology. When these paradigms are unified by the concepts of energy and information, what emerges is a radical new cosmic narrative that has implications for just about every academic discipline, from biology and psychology to quantum mechanics and cosmology.

Remember, the aim is not to reduce the higher sciences to particle physics, which besides being an impossible task turns out to be theoretically misguided. Instead, our goal is to provide a paradigm that unifies the sciences with a set of common principles while allowing space for each to study their novel phenomena somewhat independently. Discovering a high correlation between fields of science once thought to be unrelated would indicate that they inform and reinforce each other, rather than conflict with one another, as they do so flagrantly in the reductionist's conception of reality. Achieving this unification would be the unity of knowledge envisioned by E.O. Wilson, and the unity of sciences advocated by Ludwig Bertalanffy, the founder of *general systems theory*, the first formal non-reductive theory of everything. In honor of these men, let us call this unification "the Great Consilience," and hope our unifying theory of reality is up to the task. Assuming it is, what would the Great Consilience teach us that we didn't already know?

## The Self-Organizing Universe Wakes Up

By now we know that the reductionist paradigm—which represents the mainstream philosophy in physics—says in no uncertain terms that life is utterly insignificant in the cosmic scheme of things. According to this worldview, held by celebrity physicists like Neil deGrasse Tyson and Brian Greene, the universe as a whole is destined to grow increasingly disordered and devoid of pattern or meaning. But this pessimistic model of reality is being challenged as the paradigm of emergence grows increasingly popular

among the physicists, complexity scientists, and neuroscientists studying self-organizing systems at all levels of scale. What these investigations collectively reveal is that the universe seems to be evolving according to a developmental plan that is somehow encoded in the laws and constants of physics the way an organism's developmental trajectory is encoded in DNA. (This concept is explored in Chapter Thirteen.)

It turns out that the tendency for entropy to increase, described by the second law of thermodynamics, is the same "force" that drives the continual creation of organization and information in nature. This ordering tendency has been called a causal entropic force, and it constructs increasingly powerful complex adaptive systems by harnessing the energy in the environment. As biospheres emerge and evolve to better facilitate planetary energy flow, adaptive complexity gets pushed further and further from thermodynamic equilibrium by the solar gradient. Given enough time, recursive self-organization will inevitably produce all kinds of wondrous emergent phenomena with surprising new properties, like agency and mind. These emergent properties help further propagate life in the universe, so that free energy is continuously consumed (and entropy produced) at the fastest rate physically possible.

Physicist Paul Davies described this grand and majestic self-organizing process as following a "cosmic script," while Harvard educator and astrophysicist Roy Gould called it a "building plan." "Cosmic code" was the choice of words for the quantum physicist and humanitarian Heinz Pagels, Seth Lloyd's mentor and a prescient thinker. Freeman Dyson, Harold Morowitz, and Christian de Duve all held similar beliefs. These scientists, and all of those we've mentioned throughout this unfolding story, are naturalists in every regard, not to mention fierce critics of pseudoscience and superstition. If they are perceived to be mystics by today's radical reductionists for these sorts of teleologically tinged statements, it is due to the critic not understanding that nature can be inherently creative and goal directed without harboring any supernatural or mystical forces. Language gets tricky here, though, because what is more "mystical" than the cosmos naturally organizing itself into an increasingly sentient state? As our understanding of nature evolves, so must our language.

Call the proponents of natural teleology "mechanistic mystics" if you like, but that is a title that would apply to most of the great scientific minds of our time, and of history—including, though in no way limited to, Sir Arthur Eddington (unashamedly), Lord Kelvin, James Maxwell, Werner Heisenberg, Max Planck, Albert Einstein, Wolfgang Pauli, Erwin Schrödinger, and last but not least, Kurt Gödel, though you would never know that from the way these men are selectively quoted. In the final chapter we will come back to the way spiritual science history has been swept under the rug, but for now, we should know that it happened for quite understandable reasons, one being that science, especially evolutionary theory, has always been under attack from religion. This war between science and religion made it hard for non-reductionist concepts like emergence to gain credibility, though the pressing need to explain complex emergent phenomena like life and consciousness is breaking down boundaries that should have never existed in the first place.

In the coming chapters, our unifying theory of reality will bridge the gap between the physical and the mental by explaining how the realm of experience emerges from matter and material processes. This will reveal a deep connection between mind and cosmos. But before we get into the mechanisms underlying consciousness and free will, we must first become familiar with the long philosophical history of these concepts.

## What Is Consciousness?

In his famous 1974 "What Is It Like to Be a Bat?" paper, the philosopher Thomas Nagel proposed an informal but practical definition of consciousness that has been adopted by many contemporary neuroscientists because it gets the job done. Consciousness is simply "what it is like to be something." If there is something that it is like to be some system—some point of view that allows an observer to experience the world subjectively—then that system has consciousness. We all know that there is something it is like to be a human, and it seems pretty likely that there is something it is like to be a bat, but there is presumably nothing it is like to be a rock, or a river, or a grain of sand (though the philosophy known as panpsychism famously challenges this presumption).

Subjective experience begins when we wake up in the morning and ends when we fall asleep, and it picks back up again when we enter into a dream. We know for a fact that not all brain states are conscious states because awareness fades in and out, and when it's gone, we don't even know that any time has passed until consciousness resumes and we look at a clock. Experience isn't something associated with any old object; even brains must be doing a special kind of information processing for a mind to emerge.

The first clear attempt to establish consciousness as a phenomenon worthy of intellectual investigation occurred roughly four hundred years ago, and we are still arguing about it today. On one fateful night in France, the epic philosopher, mathematician, and scientist René Descartes engaged in an epistemological exercise similar to the one we went through in Chapter Six in which he began to systematically question everything he knew about existence, sending him into a downward spiral of doubt that deeply disturbed him. This method was even given a name—the "method of doubt," Descartes would call it—although the exercise is just as much about discovering truth as it is about doubting it. In his "evil genius" argument, he articulated his version of the problem of epistemology with a clever and trippy thought experiment that has inspired much sci-fi since his time.

## What Is Real?

By asking us to consider the possibility that the reality we experience is the artificial creation of some sort of evil genius with godlike powers, hell-bent on deceiving us, Descartes logically proved that we can never know with absolute certainty whether the reality we experience is in fact real or some kind of illusion. After all, dreams often appear completely real and logically consistent when we are immersed in them. The same goes for hallucinations. The conclusion is that we cannot ever be sure if we are awake in reality or inside a dream that our mind has constructed. This is a hard problem for philosophers, so let's call it the *hard problem of reality*. There just doesn't seem to be any systematic way to determine what is real and what is illusion.

The modern version of Descartes's thought experiment might say we can never know for certain whether we are in *true* reality, called *base*

*reality*, or instead inside some sort of computer simulation designed to emulate reality. This possibility was portrayed in the blockbuster movie *The Matrix*, which revives Descartes's demon in the form of a machine intelligence that programs a virtual reality so convincing it fools humans into believing the simulation is reality. The film inspired the Oxford philosopher Nick Bostrom to propose the "simulation hypothesis" two years after its 1999 release, in a provocative paper titled "Are you living in a computer simulation?" which has convinced some high-profile figures, like Elon Musk, that we are highly likely to be the digital dreams of some supercomputer.

Today, this line of thinking—questioning the nature or authenticity of reality—might sound like the superficial ponderings of a stoner, but it actually indicates that Descartes understood something quite deep and subtle about the nature of mind and cosmos: Our mental experience is only a projection, a kind of virtual representation of the real world "out there," and this mental map or model is subject to error. By thinking about the nature of reality, Descartes learned something about the nature of mind.

## I Think, Therefore I Am

Despite the skeptical and solipsistic epistemological exercise, Descartes's ultimate conclusion was that whether or not our perception of the world is accurate, the act of perception alone means that the *observer is real*, even if what he or she is observing is not. It was through thinking about thinking, a self-referential act, that led Descartes to realize that "I cannot doubt that I am doubting." Like magic, self-reference brought him out of the nihilistic nightmare he had reasoned his way into. The method of doubt had led him to the truth of his own existence.

Realizing that only a conscious entity has the ability to doubt inspired Descartes to make the simple but powerful statement "Cogito, ergo sum," which roughly translates to "I think, therefore I am." In other words, the fact that I am having an experience right now means that I—as a conscious observer—must really exist. Otherwise, who or what would be posing the question "Do I exist?" What's the lesson to be learned from Descartes's thought experiment? You can doubt almost everything about existence, but

you can't doubt that you exist. Consciousness cannot be denied or explained away. The reductionist worldview has never been able to account for the existence of subjective experience, and for this reason it fails as a model for all of reality.

While Descartes deserves much credit for establishing consciousness as a phenomenon worthy of philosophical and scientific investigation, he is also to blame for much of the confusion surrounding it. While he believed that the mind had some sort of relationship with the physical brain (he wrongly associated the pineal gland with the source of consciousness), he also argued that it was a "nonphysical" entity that operated independently from the laws of physics, or at least those that were known at the time.

And in that sense, he was not entirely wrong—understanding consciousness requires a computational framework that goes beyond classical mechanics—but what he did get wrong can be attributed to holding untested, unsupported beliefs. Being a devout Catholic, Descartes equated the mind with the Christian concept of the soul, which is ethereal and eternal. This motivated him to invent the philosophical stance known as *dualism*, which says that reality consists of two kinds of distinct substances: physical and mental, matter and spirit, body and soul, brain and mind.

Because he believed a mind is a soul, Descartes concluded that humans are the only animals that are actually conscious. No soul means no mind. Other creatures, despite appearing to have awareness, and behaving as if they experienced joy and pain, are simply automatons: fleshy machines with no inner sensation, unable to feel or suffer. He surely knew that animals like dogs have pineal glands, the suspected source of consciousness, but it appears religious dogma prevented him from applying his theory to other species. As a result, countless dogs were tortured as he experimented on them in an attempt to better understand human anatomy and physiology. This serves as a prime example of how bad philosophy and an unscientific understanding of consciousness can lead to serious ethical and moral atrocities.

Fortunately, the hard-nosed empiricists of later centuries recognized that Descartes's concept of a mind that works independently of the brain could not be squared with modern science. How could something described as "nonphysical" possibly have any effect on the physical world? This paradox is known as the *mind-body problem*. We live in a mechanistic universe,

where physical effects have physical causes, and an immaterial mind that pushes stuff around sure sounds a lot like a spirit. After the twentieth-century British philosopher Gilbert Ryle characterized Descartes's model of the mind as being like a "ghost in the machine," mind-body dualism was labeled a supernatural idea that had no place in science.

It would seem that intellectual progress was being made. However, the scientists and philosophers who rejected Descartes's dualism proposed a solution that was equally ill conceived, adding to the confusion rather than resolving it. Since Descartes had defined the mind as something inherently nonphysical, the reductionist solution was to remove it from the picture altogether.

The reductionist's alternative to Descartes's substance dualism was named *materialism*, though later it would be rebranded as *physicalism* so that it could accommodate more abstract physical concepts like energy. Materialism helped rid science of supernatural concepts like souls, but at the expense of ignoring the phenomenon of conscious experience entirely, the one thing in reality we cannot doubt, as "cogito, ergo sum" taught us. Under this stance, consciousness is considered to be an illusion, with no power to affect anything in the real world. If we are to accept the materialist picture of reality, we must accept that our experience of free will is also an illusion. This leads to all kinds of logical absurdities, but that hasn't fazed radical reductionists like German physicist and popular YouTuber Sabine Hossenfelder, who has said with complete certainty that we do not have free will, over and over, cementing it into the collective psyche. In a 2016 blog post titled "Free Will Is Dead, Let's Bury It," Hossenfelder wrote: "I wish people would stop insisting they have free will. It's terribly annoying. Insisting that free will exists is bad science, like insisting that horoscopes tell you something about the future—it's not compatible with our knowledge about nature."

Not all physicists agree. "I'm sure we have free will," David Deutsch told science journalist John Horgan in a 2018 video interview,[1] and he's in good company. Some of the world's most respected neuroscientists, like Giulio Tononi and Christof Koch, have also recently argued for the reality of free choice on a mechanistic basis.

Those who believe that our agency is an illusion contradict themselves at just about every waking moment. Ultimately they end up having to accept that the choices they are making in daily life are real at some level—otherwise, every moment would be filled with cognitive dissonance. As we can see, the materialist worldview is inherently inconsistent because the premise is fundamentally in conflict with the truth of experience. We can call materialists "dualists in denial" who acknowledge their agency in casual conversation but deny it when speaking academically because they live as though they have conscious control over their actions, despite arguing that all their actions are predetermined by physics.

For example, in a 2019 podcast episode, popular author and neuroscientist Sam Harris—who has made a career out of denying free will—and his wife Annaka Harris, author of the book *Conscious: A Brief Guide to the Fundamental Mystery of Mind* (who also denies free will), can be heard discussing whether or not they should inform their child that they have no free will. The fact that they are even having that discussion suggests that they both believe deep down that they have some choice in the matter!

Our courts are based on the idea that we are agents who are at least partially responsible for our actions (assuming cognitive health). If none of that is true, then the justice system would certainly be unfair and unjust. I'm not saying it's not—I'm just saying it's not for that reason. If it is true that none of us have any free will at any moment, we should not just be absolved from all punishment; any kind of reward would be equally undeserving. So, MVP awards, Oscar awards, Grammys, Pulitzer and Nobel prizes—none would be deserved, because none would have been earned. When it's all just particles bumping around, and that is the full story, agency and choice can only be an illusion.

In her same essay mentioned previously, "Bio from Bit," astrobiologist and theoretical physicist Sara Walker explained why many mainstream physicists feel the need to take this nonsensical position: "It is often argued the idea of information with causal power is in conflict with our current understanding of physical reality, which is described in terms of fixed laws of motion and initial conditions. If deterministic laws describe all of reality, there is no 'room at the bottom' for macro systems (such as living entitles) to do causal work."

This assumption is based on the cosmic worldview popularized by the eighteenth-century French polymath Pierre-Simon Laplace, who was inspired by Newton's laws of motion and wanted to apply them to everything in the universe, including individual atoms.

## The Logic of Laplace

In Laplace's model, the unfolding of reality is nothing more than particles of matter obeying fixed laws of motion, such that their trajectories through space and time are entirely decided at the moment the universe came into existence. So, if you could somehow know the exact state of the universe at any one point in time—if you could measure the position and velocity of every atom in the cosmos—then you could know everything about the state of the universe at any other point in time. Everything about the future would be perfectly predictable with a straightforward calculation, including all human behavior, like what you were going to do forty-two days or forty-two years from now. This model suggests we have no control over anything that happens in the world or in our lives. We do not really make decisions, because there are no choices in Laplace's universe.

*Laplace's demon*, a thought experiment created by Laplace, imagined a godlike entity that had this ability to measure the state of the whole universe and compute its future, but a modern version might feature an unimaginably powerful artificial intelligence playing the same role. If such a task were possible, it would lead to all kinds of paradoxical results. For instance, if this AI could tell you what you were going to do at some future point in time, couldn't you just do something different and defy the prediction? This paradox should be a clue that something isn't right with this picture.

Despite its shortcomings, this philosophy became the gospel of the reductionist, and Laplace's model was named *causal determinism*, or simply *determinism*, or more boldly *scientific determinism*, a name designed to portray anything that goes against the doctrine as fundamentally unscientific. Interestingly, it is also called *metaphysical determinism*, because some philosophers recognized that it was a metaphysical position, and a statement about how reality may or may not be. As we will see, there

are alternatives to determinism that are equally rigorous models of causation.

According to Laplace's model, all causation is "bottom-up," meaning the particles that all objects and organisms are made of at bottom are the physical entities that do all the pushing around. Despite the feeling that we consciously make decisions all the time, mental activity is nothing more than particles following predetermined paths. *You* don't lift your arm, a chain of events at the molecular level is what underlies the arm's movement. So, if we are to accept this position, which philosophers now call *hard determinism*, we must also accept that free will is an illusion, and a really strange one at that—an illusion we must live our life according to! Some would call that kind of illusion a delusion. If hard determinism is true, we're all delusional.

The American philosopher Jerry Fodor, one of the pioneers of cognitive science, protested this view with a quote that expresses how we all really feel about a reductionist reality deep down, if we are to take its implications seriously: "If it isn't literally true that my wanting is causally responsible for my reaching and my itching is causally responsible for my scratching . . . if none of that is literally true, then practically everything I believe about anything is false and it's the end of the world."[2]

Despite the massive influence that Laplace's vision had on science and philosophy, when quantum mechanics arrived on the scene about a century ago, there was a shift away from classical determinism toward a model of reality that was fundamentally probabilistic, reviving the idea of *indeterminism*, which says the future is not determined in the strict sense once imagined.

## The Slow Death of Determinism

The scientist that resurrected indeterminism was Werner Heisenberg, a German physicist who was awarded the Nobel Prize in 1932 for his pioneering work in quantum mechanics. His uncertainty principle demonstrated that it is impossible to have a Laplacian demon, not just in practice but also in theory, because you can never measure the precise state

of a single particle, let alone the state of every particle in the universe. Because measuring an atom's position in space affects its velocity, and measuring its velocity affects its position in space, it is simply impossible to measure a system so small without changing the value you set out to measure. Not only can we never know a particle's position and velocity with the certainty Laplace hoped for, experimental evidence has now led to the general consensus that until a quantum system is measured—an electron is a common example—it has no well-defined existence in space. Instead there is just a potential to be in one place or another with a certain probability. So, it is not only impossible for us to know where an electron will be next, it seems that *nature itself does not know*.

Around the same time that some physicists were discovering that nature is quantum at bottom, others were describing a similarly curious phenomenon known as *deterministic chaos*. In the early 1900s, physicist Edward Lorenz showed that "chaotic" systems, weather systems being his area of expertise, cannot be predicted in any perfectly precise way. This is because when you measure the initial state of the system and put that value into a computer, you have to enter a finite number of digits, and just the tiny bit of error from rounding off decimal places is enough to completely throw off the calculation's accuracy. Because chaotic systems can quickly self-amplify through positive feedback loops (remember the butterfly effect), any inaccuracy at all in the initial measurement will produce hugely inaccurate predictions of future states. The further into the future one tries to peer, the greater the prediction error grows. The precise future path of a hurricane or an organism is simply not something that can be computed by any conceivable computer.

Having already been weakened by the bizarre discoveries of quantum mechanics, deterministic chaos dealt the deathblow to Laplace's demon. Unless the demon can store an infinitely long measurement value in memory, it will not be able to predict the future. Even God wouldn't be able to predict the future perfectly. This would seem to further suggest that the future is actually being created (computed) in real time, and exactly where it will go, no one can know. While this apparent confirmation of unpredictability and indeterminism in nature sparked new discussion around the topics of free will and agency, most scientists and philosophers

ultimately concluded that indeterminism did not rescue free will, since a reality that evolves randomly seems to be equally beyond our control.

It is true that quantum indeterminacy does not provide a mechanism for free will. What it does do though, as Karl Popper recognized, is provide the necessary "wiggle room" in the causal chain that could allow room for agency or free will to intervene. The slight degree of indeterminism in nature creates "causal slack," explained physicist George Ellis, a distinguished professor of complex systems at the University of Cape Town in South Africa and a leading cosmologist. If there is some slack in the cosmic chain of cause and effect due to intrinsic randomness, then an adaptive agent with cybernetic control can act with true freedom by initiating causal chains that influence the trajectory of cosmic evolution. The fundamental fuzziness of the quantum world may seem weird, but without some degree of indeterminacy, we'd have no freedom, so maybe the architecture of reality is only as weird as it must be for conscious agents to have choice, which gives life purpose and meaning.

Despite being disproven, the nihilistic, deterministic vision of Laplace still reigns supreme in science, and that should concern all of us, because it affects all of us.

## Laplace's Demon Is Truly Evil

A philosophy that asks you to deny your own agency might sound funny, but it is really no laughing matter. As studies have shown, convincing people that they don't have free will can lead to immoral behavior. One well-known experiment found that participants who stopped believing in free will after reading a persuasive passage by Francis Crick were more likely to cheat on a subsequent test because they no longer felt personally responsible for their actions. Other studies have shown that people who have been convinced they do not have free will report increased feelings of depression or anxiety. On the other hand, belief in free will is correlated with better job performance, happier lives, and a more stable society. So, Laplace's demon, the thought experiment that birthed determinism, is truly a demon, because it is an idea that robs you of your free will by convincing you that you never had it in the first place.

If you still think it's a harmless philosophical position to believe that minds and subjective states do not exist or are illusory, this might change your mind.

In the first half of the twentieth century, the materialist stance invaded the field of psychology, which had been based on the reality of mental states up until that point. The result was *behaviorism*, an approach to understanding human nature that as a rule ignored all internal experience, like thoughts and feelings. There is no doubt it was a backlash against Freudian psychology, which was highly subjective in nature. To behaviorists, all that mattered was how a living organism behaved in response to a stimulus—the input-output mapping—not what was going on inside their mind. Remember, as reductionists and materialists, behaviorists don't believe in minds. According to this approach, there is no difference between one mental state and another unless there is some observable difference in the behavior associated with each state. And if there is no behavior, we can assume there is no active mental state.

We can immediately see the serious flaw in this way of thinking by considering the example of patients with locked-in syndrome. People suffering from the extreme form of this condition are still conscious but are completely paralyzed and unable to communicate with others. Though they show no outward signs of awareness due to brain damage, they are actually experiencing everything that is happening around them just as an unafflicted person would, but they are forever unable to let anyone know. It sounds like a nightmare, but it is a nightmare that real people are out there living. According to the behaviorist philosophy, a locked-in patient would have no inner world of experience whatsoever. Since they are behaviorally equivalent to those who are in unconscious vegetative states, they are assumed to be mentally equivalent as well. If this is the case, we should not worry about locked-in patients or their inability to communicate. Thanks to brain imaging technology and clever experimental design, we can now prove that this is not the case.

Once again, we see how an inaccurate philosophy of mind leads to serious ethical issues that we absolutely cannot ignore. Fortunately, behaviorism was a short-lived movement, as it was crushed by the cognitive revolution of the 1950s and 1960s, which is still illuminating the true nature of brain and mind today.

# Cognitive Science Saves the Day

While behaviorism did a lot to advance psychology by introducing objective measures to a field that relied almost exclusively on subjective reporting, which is often biased and inaccurate, its policy of ignoring mental states was ultimately deemed unacceptable. Inspired by the invention of the digital computer—a machine that could solve complex math problems by manipulating symbols—a new generation of scientists adopted the view that the brain was an information-processing system.

With this great insight, the field of cognitive psychology was born. According to this paradigm, the mind is real, but it is essentially an operating system—software running on the brain's hardware. While early cognitive scientists knew that the brain's wiring and overall design was radically different from that of a computer, they recognized that both systems use binary logic to compute. Recordings from electrodes implanted in the brains of various animals revealed that neurons communicate with one another through electrical signals, either by firing or not firing, thereby registering a 1 or 0. From this they concluded that mental states are created by patterns of electrical activity in the brain, such that a change in mental state is always associated with a corresponding change in the electrical brain state. Brains are still machines as the behaviorists believed, but on top of being mechanically reactive, they process and integrate information internally. Progress in understanding the brain-mind connection was being made, but slowly, mainly because techniques to record cellular activity were limited in what they could show, and they were invasive, typically requiring surgery.

But everything changed in the 1990s, when the brain imaging technology known as fMRI (functional magnetic resonance imaging) allowed cognitive scientists to noninvasively peer inside the brain to see exactly what the mind was doing in real time. This new tool, enhanced by the statistical analysis techniques developed by Karl Friston, would invigorate consciousness research and bring it out of the fringe and into the academic mainstream. For the first time, the neural activity underlying states of conscious experience could be mapped out, and in impressive detail.

Francis Crick, working in collaboration with the promising young neuroscientist Christof Koch, set the agenda. Progress in understanding the mind could be made by discovering which parts of the brain are active only during states of consciousness. These brain regions are referred to as the *neural correlates of consciousness*, and over the next three decades, these correlates would be more or less identified. While this increased our understanding of awareness by leaps and bounds, much of the mystery of consciousness remained. What was it about the activation of those particular neurons that creates the mysterious subjective states of conscious experience? And, why should information processing in the brain *feel* like anything? After all, mindless machines like calculators and thermostats also process information.

After Crick and Koch's approach seemed to hit a dead end, a new theory of consciousness emerged on the scene called *integrated information theory*, proposed by the Italian American psychiatrist Giulio Tononi in 2004, whose genius idea reinvigorated consciousness research and stirred excitement that the hard problem of consciousness would finally be solved by science.

## Consciousness as Integrated Information

Why does the global synchronous firing of certain networks of neurons create a subjective field of experience, according to Tononi's theory? Because the brain is not just *processing* information, but *integrating* information into a unified whole. Just like a computer, the brain stores and processes information, but it is how that information is shared or distributed throughout the brain network that gives rise to our rich and vivid conscious experience.

Let's think about the act of observing a sunset. Thanks to functional imaging, we know that there are a number of different brain regions active during such an event. Each region processes information about different features of the event separately. There's a region in the visual cortex that processes the form and color of the yellow and orange sun rays against the clouds. Auditory areas in the temporal lobe are fed information about the sound of the wind rushing past you as you stare off into the horizon. The

rushing wind against your skin also generates patterns of electrical signals in the somatosensory cortex that create a sense of touch. Inside the brain, there are many different things going on in distant places. Yet, somehow we perceive it all as one unified conscious experience. Thanks to integrated information theory, we are now beginning to understand how the perceptual binding of independently processed features happens mathematically.

Each possible experience you could conceivably have corresponds to a very specific mathematical structure with a unique information geometry. This structure represents the shape and character of a subjective experience, while the intensity or richness of that conscious experience corresponds to a quantity called *phi*. In Chapter Eight, we learned that phi is a good measure of computational complexity because it quantifies a cognitive system's integration and differentiation. Phi also measures what Tononi calls "cause-effect power," so it would also seem to quantify the degree of agency an agent possesses—this type of causal power is a bit different from the kind we have been discussing, which is more about the magnitude of influence an agent can have on its environment, where phi measures the causal power a system has *over itself*.

It might sound loopy to think about how much influence a system has on itself, but this only means that phi is a measure of how much the present state of a system physically influences its future state. Despite being a system composed at bottom of components (molecules) whose behavior has a random or probabilistic character, we as functional wholes are deterministic machines that move with precision and purpose. This kind of purposeful or teleological behavior is only possible if the system constrains itself in a way that preserves the encoded information and transmits it to the future. In terms of Shannon's information theory, phi is a measure of mutual information, or shared information, between the system's present and future state (or the past and present state). While there are important distinctions between the two concepts of causal power, this measure of intrinsic information would presumably be correlated with how much influence an adaptive system can have on the world around it. A similar measure has been invented by Santa Fe Institute president David Krakauer, outlined in a framework called the *information theory of individuality*.

Koch, Tononi's most high-profile collaborator, describes phi as a measure of synergy, or the degree to which a system is "more than the sum

of its parts." Phi has also been characterized by Koch as a measure of uncertainty reduction, so phi must in some way reflect the amount of knowledge the system has about its environment. As also explained in Chapter Eight, phi can be seen as an entropy measure, or a measure of the number of mental states that are potentially accessible to the agent. So, a human brain would have a much higher phi than a cat, and a cat's phi would be higher than that of a rat, and we might guess that the rat's would be a little higher than a bat's.

It only makes sense that the degree of conscious experience at any moment, how rich it is perceptually, would correspond to the number of states that *could be* experienced by the system, which is the information entropy. Think about it this way: A system with two states can only know black and white, one or zero, good or bad, etc., so a conscious observation would yield only one bit of information. If a sentient system can only experience a binary world, then there's no sophistication or richness to the reality that is being experienced. It is the spectrum of possible experiences we could have but are not having at the moment that make the present experience what it is.

So, to use a metaphor, if a six-sided die were sentient, it would have a richer experience of the world than a coin, because it can experience three times as many states. One cannot know extreme beauty or pleasure without also knowing the complete opposite, or at least being able to conceive of it. Beyond having a kind of Zen aesthetic that might offer some spiritual comfort, this concept of phi can inform our moral philosophy. The theory means that a system as simple as shrimp, certainly a mussel, does not have the capacity to suffer nearly as much as a mammal. And when you drive your car and an insect splatters against the windshield, you don't have to feel all that bad about that kind of murder. Conversely, animals shown to be highly intelligent, like a dog or octopus, would suffer greatly if they are being tortured or killed with no anesthetic. In fact, they presumably suffer in much the same way a small child would. On a cheerier note, it would also mean that as humans evolve and adapt to the growing number of existential challenges that humanity faces, our world model will be mapping more and more of reality, and as our spectrum of possible experiences increases, so should the intensity with which we feel them. The future will be a really feely place for whoever is still around to feel it.

It might sound like integrated information theory is the answer to the mind-body problem, but before you go thinking we've cracked the case, you should know that this theory—at least in present form, since it is always evolving—leads to a philosophical conclusion that is potentially just as problematic as either the materialist or dualist picture.

## Integrated Information Theory's Panpsychism Problem

Taken to its logical conclusions, integrated information theory says that any system that has any amount of phi at all should possess some degree of conscious experience and should have some degree of causal power. This aspect of the theory is reminiscent of an old Eastern philosophy called *panpsychism*, which says that all matter is imbued with mind, whether that matter is part of a living creature or not. Of course, sandpiles or rocks as a whole wouldn't be conscious under integrated information theory—as traditional panpsychism posits—because the macroscopic system is not integrating any information. But its most basic components, the atoms that comprise it, would be "minimally aware," according to fundamentalist versions of integrated information theory.

Koch, a leading proponent of the theory, is adamant about this belief, saying "Any system that possesses some nonzero amount of integrated information experiences something. Let me repeat: any system that has even one bit of integrated information has a very minute conscious experience." This obviously has some very strange philosophical consequences, but Koch embraces them: "Even simple matter has a modicum of Φ [integrated information]. Protons and neutrons consist of a triad of quarks that are never observed in isolation. They constitute an infinitesimal integrated system."[3]

A contradiction lives inside those who believe panpsychism: they will admit that they lose consciousness between the waking state and the dreaming state, when the brain is still processing information, yet they want to attribute consciousness to an atom, or a thermostat. It is not wrong because it sounds absurd—nature doesn't care about conforming to our

common sense. It is wrong because it is equivalent to a magical explanation. The mind is a mental simulation of the world experienced by an observer, and a coherent simulation requires sophisticated computational machinery. People who imagine that an elementary particle is conscious are assuming that subjective experience is a fairly easy thing for matter to produce.

While poetic meta-naturalism says it is true that the universe can be described in entirely computational terms—meaning there is "mind" in the form of information processing throughout nature—it is not true that there is a conscious observer in all physical systems, even if there is some information integration. For consciousness to arise, which is required for free will, a system must not only model its environment, it must also *model itself*.

The unifying theory of reality will answer the seemingly unanswerable questions posed by the mind-body problem, and in doing so, will illuminate not just the nature of mind but also the nature of reality. By addressing Australian philosopher David Chalmers's *hard problem of consciousness*—the problem of how certain material systems acquire the seemingly immaterial property of *experience*—we will simultaneously solve the *hard problem of reality*, which asks whether reality is all physical, all mental, some combination of the two, or something else altogether. It is "hard," because understanding it would entail comprehending the ultimate nature of reality, and most philosophers would consider this an impossible task. But with paradigm shifts, the impossible often becomes possible. The universe can be understood best, and life can be understood best, not as a thing, but as a process.

# 11

# Causal Emergence and Free Will

---

> As we move up in scale from subatomic particles, to atoms, to molecules, to life and mind, the universe appears to become more and more complex, and it seems clear as complexity increases entirely new properties emerge. The question of interest is whether these emergent properties can themselves have causal consequences, allowing agents to have an active say in achieving their own goals.
>
> **—Sara Walker**

---

Now we are nearing the point in the story where we can begin to appreciate some of the strange and beautiful philosophical implications of the new cosmic narrative that is emerging, and soon we will see why it is indeed "romantic," as the title of this book claims. The reader hoping to understand reality in the most honest and pure sense might object to any romanticized characterizations of the universe on the grounds that such interpretations unnecessarily anthropomorphize nature and muddy up the mechanistic picture. This is a commendable concern, but we have seen that it is misguided.

The proper use of teleological language and metaphor is not just helpful—it is the only way to accurately articulate the causal structure of a universe that organizes itself through loops and levels. As we gain a deeper understanding of how the abstract words we invented to describe poorly understood emergent phenomena actually map onto newly discovered mechanisms and mathematics, the biggest unsolved mysteries of science begin to dissolve one by one.

In this chapter, we will see that the mind-body problem is no longer a problem once we more precisely define terms like mind, agency, and free will, and when we articulate the mechanisms that make each phenomenon distinct. Doing so will reveal that conscious creatures really can control their own destiny *by selecting the future that is consistent with their long-term goals from a menu of possible options*. Computational neuroscientists and philosophers call these potential futures *counterfactuals*, and when we become aware that they exist, we must then choose the particular path we take out of the space of possibilities. It is in this act where we will find the freedom in "free will," which is certainly something that humans possess, just not all the time, and it is not quite as free as we might like!

## The Many Meanings of "Mind"

The word "mind" is often used to describe a physical system specifically with consciousness or subjective experience, while other times it is used vaguely to describe any biological or cognitive computational process, regardless of whether or not the information-processing system actually experiences the world qualitatively, from a first-person perspective. "Mind" is also used in senses like the "collective mind," when referring to societies, or humanity as a whole, again where it is not known whether there is a subjective experience associated with the adaptive system being described.

If all that weren't confusing enough, we have this concept of the *conscious mind* and also the concept of the *unconscious mind*. To the people who use "mind" specifically to mean consciousness, the term "unconscious mind" will appear to be an oxymoron, as the two words directly contradict each other. Of course, once we understand that some people use the word "mind" to refer to all forms of cognition, then we can see how "unconscious

mind" is not a true contradiction, but a nuance that distinguishes between unconscious and conscious cognition, though it is a confusing choice of words.

It is not that any of these uses of the word is fundamentally incorrect—all these examples refer to information processing by an adaptive system, which is at the heart of mental phenomena. The problem is that the important distinction between conscious and nonconscious cognitive processing gets lost or obscured when we use the words "mind" or "consciousness" all willy-nilly. While it may seem pedantic to worry so much about semantics, it is this distinction that will make sense of free will.

We must not confuse basic agency (being alive) with consciousness (being aware), because you can have agency without consciousness. Free will, on the other hand, requires consciousness, though you can be conscious without having free will. Our unifying theory of reality will show us where and why each property appears as a result of cosmic evolution, and how the emergence of new levels of agency is tied to the larger process of cosmic awakening.

We are now beginning to see why the "meta" in poetic meta-naturalism is a necessary alteration of Sean Carroll's poetic naturalism. Carroll's term depicts nature as only this mechanical thing, implicitly denying the ontological reality of experience and the causal power of consciousness, which we will see are very real indeed. We are logically justified in considering them to be real causal phenomena because it is impossible to understand the dynamics of cosmic evolution without some direct reference to them.

"Meta" denotes that the causal structure of reality has levels, with new ones always coming into existence through higher order phase transitions known as metasystem transitions. The evolutionary process is poetic because it is inherently creative, and each new functional pattern that emerges at a higher level is a "rhyme" on the level before it. In poetic meta-naturalism's model of reality, the mind is a multilevel controller, and free will is a higher level of control than basic biological agency. Once we understand the hierarchical structure of life, mind, and cosmos, we begin to see that individual freedom and cosmic destiny are not incompatible notions.

Levels of agency might sound abstract, but we are about to see that this has important practical implications, from ethics to mental health.

## Levels of Agent Causation

Despite what seemed like an impossibility from the perspective of the reductionist model, the ability of a biological system to act as a causal force in a physical world has long been suspected by philosophers and implied by common sense. Our conscious experience of lifting our arm seems to tell us that at times we are agents who can exercise voluntary control—or volition. Agency, volition, will, intention—they all seem to be getting at the same general thing, and that is goal-directed action, which is not random behavior but movement guided by a purpose—or a teleology. We now know that this teleology is not the result of a mystical force but a natural product of the adaptive information encoded in the machinery of life. But is all goal-oriented movement equivalent?

We don't think much about raising our arm when we feel an itch on our head that needs to be scratched, but other times we do. When you decide to raise your hand as a deliberate exercise of free will, then the act is a conscious one and not a purely reflexive or instinctual one. You seem to be the author of the action in one instance but not so much the other.

When the arm raises without conscious thought, it is because that behavior is preprogrammed and does not require what cognitive psychologists call *effortful control*. Effortful control, also known as *cognitive control*, or *executive control*, refers to the ability to regulate emotion, behavior, and cognition. This type of high-level control seems to require consciousness, at least typically. Conscious representation of conflicting information may also allow an agent to make high-level decisions when faced with challenges that might not have a preprogrammed solution because they may have never been encountered by the agent before. By simulating various scenarios in the "mind's eye," we can choose the optimal decision and behavioral response for a novel situation.

So, it seems there is an important distinction to be made. Some *purposeful-looking* behavior is executed with no conscious effort. It just happens. You feel an itch and you scratch. No focused thought necessary.

Already, we can see that there seem to be levels of agency. All agent causation is not created equal, because higher levels of control emerge that give the agent new degrees of freedom. An agent with human-level cognition can use their conscious mind to imagine novel futures and bring those imagined futures into existence, like when someone has an idea for a new movie, business, or scientific theory that changes the world.

The mechanism of agent causation was given multiple names in the twentieth century by philosophers and cognitive scientists, including *downward causation, top-down causation*, or—when considering organisms with nervous systems—*mental causation*. We already see the various levels reflected naturally in the language that has emerged, because mental causation presumably wouldn't apply to a creature with no brain, like a bacterium.

However, because the general notion of agents with causal power couldn't be reconciled with the deterministic model made famous by Laplace, the idea of top-down causation in any form was largely dismissed by the scientific community. Recall from the previous chapter that in Laplace's universe, all causation is "bottom-up," meaning the motions of the particles a system is made of at the bottom fully determine the behavior of the high-level system, like an agent. Conversely, top-down causation refers to a higher level system having physical influence over the dynamics of its low-level components.

Even Daniel Dennett, whose career has really been about explaining how natural selection generates the biological information that instills adaptive systems with causal power, dismissed agent causation as a "fuzzy philosophical concept" with no firm basis in mechanism. This claim was made in a 2014 exchange on free will with philosopher and neuroscientist Sam Harris, who denies free will, even though Dennett himself argues for the reality of free will! But how can an agent have free will if it does not have causal power?

We call this self-contradicting position *compatibilism*, and because it tries to reconcile hard determinism with free will, it is also known as "trying to have your cake and eat it too." But at least the half-baked idea called compatibilism is half right—those who believe in hard determinism *and* that free will is an illusion are wrong on all fronts. There are no awards for consistency in one's wrongness, except maybe the "Wolfgang Pauli

award" for being *precisely* wrong rather than "not even wrong," an insult that is basically fighting words among physicists. Dennett, not the one to be timid or soft with words, said:

> Some have gone so far as to posit an otherwise unknown (and almost entirely unanalyzable) phenomenon called *agent causation*, in which free choices are caused somehow by an agent, but not by any event in the agent's history. One exponent of this position, Roderick Chisholm, candidly acknowledged that on this view every free choice is "a little miracle"—which makes it clear enough why this is a school of thought endorsed primarily by deeply religious philosophers and shunned by almost everyone else.[1]

Today that is far from true, and the theoretical work on top-down causation by physicists like Sara Walker and George Ellis, along with neuroscientists like Giulio Tononi and Erik Hoel, is bringing in some heavy artillery—new mathematics like causal calculus and Bayesian statistics—to aid in the effort of demystifying both agent and mental causation.

Ideas about agency and causation go back to Aristotle, although scientists began modernizing these quasi-mystical notions in the nineteenth century. These scientists include the great psychologist William James, as well as the British emergentists, like C. Lloyd Morgan, who actually recognized distinct levels of agent causation. But the terminology *downward causation* and *top-down causation* didn't surface until around the 1960s, first with the Nobel laureate Roger Sperry, the neuropsychologist famous for his split-brain studies with neuroscientist Michael Gazzaniga. Donald Campbell, who developed the concept of evolutionary epistemology, followed suit and influenced Popper to adopt the idea as well. Sperry described in detail what he called "emergent downward causal control," which Campbell rebranded as downward causation, and top-down causation just became another way to talk about the same thing. *Causal emergence* is a twenty-first-century buzzword used to describe the emergence of agent causation, and it is stirring new excitement around a very old problem.

Toward the end of his career, in defiance of materialism, Sperry began calling himself a mentalist, but one should not take this to mean he became

a Descartes-like dualist. Much to the contrary: Sperry was proposing a form of multilevel monism—an extension of physicalism to be exact—that considered mind something ontologically new and real in the universe, but equally physical. Whether you consider this philosophy of Sperry's to be a weaker form of dualism called *property dualism*, or an extended physicalism called *multilevel monism*, is a bit of a semantics game. Unlike Descartes's substance dualism, which says there are two fundamental substances, matter and spirit, property dualism says there's only one physical substance but it can have dual properties, physical and mental, which somehow interact causally. The models proposed by these two seemingly opposed philosophical stances—property dualism and multilevel monism—look basically identical. However, multilevel monism is a more flexible philosophy because it allows room for the emergence of new phenomena with causal power. As you can see from the following quote from Sperry, there's a lot of nuance required to make sense of matters of mind and why mind matters, but once you understand the multilevel nature of agent causation, many paradoxes solve themselves. In his article, "Mind-Brain Interaction: Mentalism, Yes; Dualism, No," published by the journal *Neuroscience* in 1980, he said the following:

> By our current mind-brain theory, monism has to include subjective mental properties as causal realities. This is not the case with physicalism or materialism which are the understood antitheses of mentalism, and have traditionally excluded mental phenomena as causal constructs. In calling myself a "mentalist," I hold subjective mental phenomena to be primary, causally potent realities as they are experienced subjectively, different from, more than, and not reducible to their physicochemical elements. At the same time, I define this position and the mind-brain theory on which it is based as monistic and see it as a major deterrent to dualism.

The term *downward causation* is often used ambiguously to mean the general type of behavior that living agents engage in whenever they exert an influence on the physical world, whether or not they do so consciously. Downward causation has even been used to describe the way the dynamical flow patterns of ordinary dissipative structures constrain the chaotic motion

of the individual molecules that make up the dynamical system, funneling them along ordered paths of least resistance. Unfortunately, this generalization obscures the distinction that Morgan was trying to articulate. That is, there are levels of causation and control, which form a hierarchy, such that the mental level associated with conscious control resides above the biological level associated with automatic behavioral responses. Basic agency emerges when life arises, before consciousness appears, so there is both nonconscious and conscious top-down causation. Conflating the two types of causation, which we may respectively call *agent causation* and *mental causation*, has caused much confusion in the free-will debate. To understand what free will means, we must first disentangle all the ambiguities that arise from an unsophisticated understanding of the underlying mechanisms of causation.

In the last decade, and especially in the last few years, the views of Sperry, Campbell, and Popper have found empirical support from cutting-edge research by neuroscientists with an understanding of information theory, computational modeling, and statistical mechanics. In understanding adaptive systems as hierarchical control units that pursue intrinsic thermodynamic survival goals, we are demystifying and mathematically formalizing the multiple mechanisms of top-down causation. While this type of causation has historically been associated with conscious beings like humans, the basic mechanism actually precedes brains, and presumably, minds.

## Unconscious Agency: The First Level of Control

According to Paul Davies and Sara Walker, the capacity for top-down causation emerges when the information stored in molecules begins telling the molecules what to do. We call this event *the origin of life*. At this point, the chemical system, now technically an adaptive or biological system, can begin to "push matter around." They refer to this ability as *informational control*, and it is this property that can explain notions central to life, such as autonomy, self-regulation, and goal-driven behavior. In Davies's words, top-down causation through informational control gives life the "ability to

harness chemical reactions to enact a pre-programmed agenda, rather than being a slave to those reactions."[2]

The emergence of top-down causation is an example of what philosophers call *strong emergence*, because there is a fundamental change in the kind of causality we see in the world.

With the phase transition to life, an event Davies and Walker call *algorithmic takeover*, physical systems became freed from the fixed trajectories that govern the movement of inanimate macroscopic systems and make them predictable from simple laws of motion. Animate systems can climb uphill, going against the force of gravity. Goal-directed behavior is seen in even the simplest conceivable organism, which suggests that cognition, in the most general sense, emerged when life began. The origin of life, wherever it occurred first, was the moment when information gained causal power in the universe. Inanimate systems may be able to be described *in terms* of information, but information becomes meaningful to reality when it can make a difference to an agent. Much of this book has been devoted to explaining the details of how computational agents with causal power emerge from a dissipative Darwinian process that represents the earliest form of evolution.

Once algorithmic takeover enables informational control, causation is no longer strictly bottom-up—there is now also top-down causal influence. The emergence of life represents the transition in a chemical system's structural and functional architecture that enabled the capacity for cybernetic control.

Top-down causation is no longer a vague and mysterious concept of philosophy but a mathematically formalized mechanism that describes how a system with control (like an organism or a brain) can steer itself toward goals, thereby influencing the paths of its low-level components (the molecules that make up the organism or brain). This means that living organisms are not simply slaves to the laws of physics and fundamental forces. As adaptive systems with intent, we create our futures in accordance with physical law but are not subservient to it.

As already stated, quantum indeterminacy does not provide a mechanism for free will, but it does provide the "causal slack" that was not afforded by Laplace's model of reality. It is the *combination* of top-down

causation and the probabilistic or stochastic (random) nature of reality stemming from quantum indeterminacy and deterministic chaos that explains exactly how living systems can function autonomously and affect physical reality through intentional action. We can actually redefine life as a system *with the ability to initiate new causal chains*. Causal chains are chains of cause and effect that result from an agent's actions on its environment.

According to the British computational biologist Denis Noble, it is an adaptive system's ability to harness and amplify stochastic fluctuations from chaos or indeterminacy that gives rise to the control we see with living organisms. Cybernetic systems use feedback loops to remain in a steady far-from-equilibrium attractor, so the same kind of nonlinear feedback loops that can create a butterfly effect are utilized by adaptive systems to initiate new causal chains. The bottom-up flow of causation is never disrupted; it is just harnessed and directed toward a goal. Although consciousness has not yet entered the picture, we already get a clue that the mind is not an immaterial substance like Descartes imagined but a control unit that has been engineered by natural selection to extract energy from the environment.

It is important to note that this information-based explanation of agentlevel control is not just a "high level description" of a purely bottom-up causal process that we use out of convenience, as implied by some versions of the philosophical principle known as *supervenience*. Supervenience simply means that high-level descriptions of cognitive phenomena—psychological descriptions that refer to processes like conscious decision making—must correspond to physical interactions happening at the level of atoms in the brain. Otherwise we'd have substance dualism, which is a mystical notion.

In the reductionist's model of supervenience, the microscale physics fully determine the mental processes that take place in brains, which suggests that the experience of making decisions is an illusion. If this is true, it would mean that neither mind nor information can be true sources of causality, as all the work is being done by the particles pushing each other around at the bottom level. However, new theoretical work suggests that this is provably false.

Giulio Tononi and Erik Hoel have taken up the task of mathematizing the emergence of causal power using formalisms adapted from philosopher and computer scientist Judea Pearl's causal calculus, which also provided the mathematical foundations for integrated information theory. In an award-winning essay written in 2016 titled "Agent Above, Atom Below: How Agents Causally Emerge from Their Underlying Microphysics," Hoel laid out his theory of *causal emergence*.

## The Macrostate Is What Matters

Hoel has shown that a system's dynamical evolution, meaning its lawful transition from one state to another, can be analyzed at different levels of scale to determine whether the system is an agent with causal power or an inanimate system with no control over its future. A microscale analysis would model the system at the level of particle collisions while a macroscale analysis might model the patterns of information flow between brain regions. If a system's evolution can be predicted more accurately and reliably by looking at the causal interactions at the macroscopic level, then there must be information at work in the system. The measure of "extra causal work" that can be attributed to the macroscale—the whole that has quantifiably become more than the sum of its parts—would indicate "causal emergence," which is to say that an agent with causal power has emerged.

According to Hoel, who ultimately wants to understand the connection between consciousness and causation, the causal power is located not in the microphysics but at the macro level, because information is encoded in a pattern of organization, which is multiply realizable. As we learned in Chapter Nine, multiple realizability means the same functional pattern can be realized on different substrates. Since many different microstates (unique molecular arrangements) can realize an equivalent global pattern, or macrostate, the details of the microstate are causally irrelevant. It is the pattern that pushes around particles, not the other way around.

In an article titled "Escaping Flatland: When Determinism Falls, So Does Reductionism," neurogeneticist Kevin Mitchell, author of *Innate: How the Wiring of Our Brains Shapes Who We* Are, explains that when it comes to causality, the macrostate is what matters:

> The macroscopic state as a whole does depend on some particular microstate, of course, but there may be a set of such microstates that corresponds to the same macrostate. And a different set of microstates that corresponds to a different macrostate. If the evolution of the system depends on those coarse-grained macrostates (rather than on the precise details at the lower level), then this raises something truly interesting—the idea that information can have causal power in a hierarchical system, and, more generally, in the universe.

Sara Walker has similarly subtle but grand statements about the causal power of macrostates: "Theories and theorizers can potentially influence dynamics at lower levels. Macrostates may not just be predictors, viz the informational properties of emergent agents, but could also play a role in the *causal structure* of reality."[3]

Now when we ask, "How could it be possible for mind to push things around?" let's replace the word "mind" with "information" and see if the mind-body paradox remains. How does information push particles around? That is a question we are beginning to be able to answer with mathematical and mechanistic precision. Hoel's work on causal emergence inspired further quantitative investigation into the phenomenon of top-down causation by some rising stars of neuroscience, like psychedelics researcher Robin Carhart-Harris and consciousness researcher Anil Seth.

A paper they coauthored with Imperial College London postdoctoral researcher Fernando Rosas, titled "Reconciling Emergences: An Information-Theoretic Approach to Identify Causal Emergence in Multivariate Data," shows that there is a mathematics, aside from Pearl's causal calculus, that describes the causal dynamics and statistical mechanics of different kinds of higher-level causation in terms of information flows. The paper identifies a new phenomenon the authors call *causal decoupling*:

> In addition to downward causation, causal decoupling takes place when collective properties have irreducible causal power over other collective properties. Interestingly, a feature that exhibits pure causal decoupling can be thought of as having "a life of its own," a sort of

*statistical ghost*, that perpetuates itself over time without any individual part of the system influencing or being influenced by it.

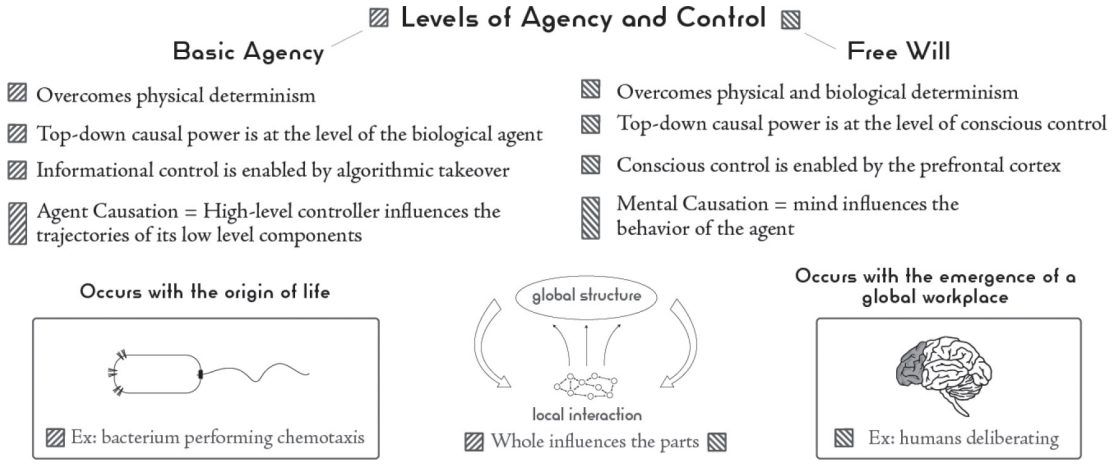
Do coordinated swirls of activity and information in the brain create a literal ghost in the machine? If so, perhaps Descartes wasn't so wrong. Suddenly the mind-body problem doesn't seem quite as mysterious, and the hard problem doesn't seem all that hard. Remember, IBM's Rolf Landauer proved in the twentieth century that "information is physical," and once we accept this fact, the dilemma of dualism dissolves away, or at least a big chunk of it. The mind can influence the physical world in a paradox-free manner because thoughts are just instances of information in action.

Although agency or control is a property present in simple life, this does not seem to be the same as free will. The bacterium performing chemotaxis will predictably swim toward chemical food and away from toxins. So, if the behavior is predictable to us, maybe not perfectly predictable in the way Laplace imagined but statistically predictable with high certainty, then how much free choice does the bacterium have to do otherwise when it senses a chemical gradient? It will typically behave as expected, and it will do so without any hesitation, so there are no behavioral indicators of deliberative thought or contemplation.

We can certainly call it purposeful or teleological behavior because it is goal oriented or end directed, as Aristotle characterized it over two millennia ago, but the agent does not seem to have any conscious awareness. Why? Because the simple organism with no brain does not respond in an adaptive way to any variables that were not present in the environment over the evolutionary history that built up the species' internal model. It "knows" only what the genome has encoded, and it is "blind" to everything else. This would suggest that the adaptive system is an automaton, mindlessly following some sort of wired-in cognitive-behavioral program.

Although agent causation doesn't seem to get us quite to free will, it does get us one step closer to it. Instinctual behavior may not be totally free, but it is not determined in the strict sense envisioned by Laplace. That is to say, top-down causation—when information starts steering macroscopic systems away from the trajectories predicted by classical mechanics—frees adaptive agents from *physical* determinism. But organisms with downward

causal power still seem to be shackled by the constraints of *biological* determinism, and for organisms with brains, *neurobiological* determinism. Automatic behavior does not seem very "free."


Levels of Agency and Control
Basic Agency
Overcomes physical determinism
Top-down causal power is at the level of the biological agent
Informational control is enabled by algorithmic takeover
Agent Causation = High-level controller influences the trajectories of its low level components
Free Will
Overcomes physical and biological determinism
Top-down causal power is at the level of conscious control
Conscious control is enabled by the prefrontal cortex
Mental Causation = mind influences the behavior of the agent
Occurs with the origin of life
Ex: bacterium performing chemotaxis
global structure
local interaction
Whole influences the parts
Occurs with the emergence of a global workplace
Ex: humans deliberating

*The free will topics are discussed in the next chapter.*

As we've established, even the simplest organisms have a primitive form of cognition, which allows them to sense and respond to the surroundings, but we should not mistake agency for consciousness. Agency only requires that a system have a world model—an internally embedded statistical mapping that represents the biologically relevant regularities in the environment. But the mere presence of a world model doesn't mean that the contents of the model are being experienced by an observer. Subjective experience is always witnessed by a "self," which forms the seat of a first-person perspective. So, it seems that progress in solving the hard problem of consciousness can be made by identifying the mechanism that brings about an observer with a point of view. What, then, is the source of the self?

According to the cognitive philosopher Douglas Hofstadter, for a point of view to emerge, the agent's world model must *model itself*. We call this example of self-reference *self-modeling*, and for reasons that will soon be explained, it is an ability that would seem to require a brain. If Hofstadter's model of minds as models with self-models is correct—which he described in qualitative detail in his 2007 book *I Am a Strange Loop*—we finally have a relatively precise answer for when consciousness comes into the picture as a result of cosmic evolution. Mind, in the full sense of the word, emerges

with brain. Self-modeling as the source of self might sound too obvious, but sometimes the answers to problems that appear difficult are staring us in the face.

Want proof that you are an observer with a model of the world and of yourself inside your head? Let's do the exercise we did in Chapter Six. Close your eyes and try to envision yourself in a room. Zoom out to see your house and your neighborhood from a bird's-eye view. Zoom out and see the planet Earth from space. Now imagine a friend sitting next to you in your room. You have not only modeled the world in your brain, you have modeled other modelers, and those familiar agents come to life in your dreams, complete with personalities and trademark quirks!

These models of other minds were built up from accumulated experience with those people. While we are all unique individuals, we are also in many ways the sum total of all the minds we have modeled, and without knowing we consult with these other minds when making difficult decisions. For example, you may call on a famous musician you admire when writing a melody for a song, and you may call on a role model when trying to make a decision about your future, like a parent or an influential teacher you once had, without any real conscious thought or reflection.

These agents need not be real agents that we've encountered in reality either. The popular saying "What would Jesus do?" reveals that Christians who ask this question have modeled their deity and look to that model for advice, support, and existential insight. The same goes for every other religion, and when we realize this, we see why there can be practical benefits to modeling abstract entities that have useful characteristics or properties. This general process has been called "thinking through other minds" by the same researchers who have applied the Bayesian brain hypothesis/free-energy principle to high-level hierarchical systems beyond brains, like ecosystems and societies.

Friston and colleagues Maxwell Ramstead, director of research at the Spatial Web Foundation in California, and cognitive scientist Samuel Veissière at McGill University in Montreal have argued that the process of inferring other agents' beliefs and expectations about the world informs us on how to behave in social situations and brings stability to society by coordinating collective behavior.

Now imagine yourself in your room, and try to stare at yourself from above. How do you look? This is proof that you've modeled yourself as well. Pretty loopy, right? Now ask yourself, how accurate is your model of you?

## How Brains Bring About Observers

Cogito, ergo sum (I think, therefore I am, as mentioned in the previous chapter) tells us the self is an undeniable fact of existence, but how does an observer emerge exactly, and why might such a process require a brain?

In Part Two, we learned how a model of the external world gets encoded in an organism in detail. Before brains emerged through evolution, there was only phylogenetic learning, which is "belief updating" that occurs at the level of the population as the genome of the species evolves over long periods of time. This is as a result of competition between organisms, with natural selection filtering out the inaccurate models and dysfunctional designs. Recall that evolution is a process of *Bayesian model selection*. However, the collective genome updates much more slowly than brains, which encode the consequences of causal actions in real time, thanks to *synaptic plasticity*. Plasticity refers to the brain's ability to continuously form connections that encode representations of the patterns in the world that the organism has experienced firsthand.

An agent acts on its environment, observes the change in that environment, updates its model by encoding this change in memory, and through iterations of this process, the organism's world model *creates a data variable for itself*. Since brainless organisms can't encode the causal consequences of their actions, they lack a variable for themselves and any real understanding of themselves as a causal actor. Because they lack self-modeling capacity, brainless organisms are presumably what philosophers call "zombies," which are systems that do not experience the world but appear to move with conscious intent.

It is self-reference through self-modeling that brings subjective experience into the world by creating an observer out of thin air. The self is the entity that is the "I" we are all referring to when we say "I feel tired," or "I love you," or "I love this book I'm reading." An observer is born when

there is “something it is like” to be that thing. We have already explained that there is nothing it is like to be a rock, despite what some panpsychist philosophers are preaching today. It is when a world modeler models itself that the “light comes on” and first-person experience emerges.

To be clear, basic consciousness is not the same as self-consciousness or self-awareness, which is required for consciousness to have causal power. It is easy to imagine having a perceptual experience without any self-reflective thought. “Spacing out” is one example of having an experience and even a train of thought without being aware of the fact that one is having a train of thought. When caught spacing out, we often have no recollection of what we were spacing out about.

So, observers may not be aware that they are observers. They may just be experiencing the world without any real contemplation about the nature of nature or the nature of oneself. This suggests that consciousness likely starts off epiphenomenal, which means at first it is simply a secondary effect or useless byproduct of the brain with no causal power over anything. The most basic agent with a brain is no longer a zombie (an agent without mental experience), but it still does not have *conscious control* over its responses. It is a passive observer and a prisoner to its programming. We should not think of this as an unpleasant state of existence though. Koch explains what he thinks it might be like to be a bee, an organism that is likely to be conscious but unlikely to be self-aware:

> They do very complicated things. We know that individual bees can fly mazes. They can remember scents. They can return to a distant flower. In fact, they can communicate with each other, through a dance, about the location and quality of a distant food source . . . So yes, I do believe it feels like something to be a honey bee. It probably feels very good to be dancing in the sunlight and to drink nectar and carry it back to their hive.[4]

So, it seems that being conscious is not enough for what we call “free will”—that comes later. The conscious mind only gains causal power when there is an additional level of self-modeling. We’ve explained a very simple reason why brains support self-modeling; they allow an adaptive system to encode the causal consequences of their own actions, but this fact alone is

not a satisfactory explanation for how the brain creates the conscious mind—or how the mind exercises free will. For that, we must look to the modern theories of consciousness that neuroscience has to offer. What are the neural and computational correlates of self-modeling? Again, we come back to the notion of loops and levels, and to understand their relevance to consciousness and free will, we have to get Gödelian, which means we must explore the causal magic enabled by self-reference. This investigation will reveal the connection between the emergence of higher levels of cybernetic control and cosmic self-organization, an inevitable evolutionary process that curiously depends on intelligent agents who have the power to choose their future path.

# 12

# The Strange Loop That Spawns the Self

---

> Either mathematics is too big for the human mind or the human mind is more than a machine.
>
> **—Kurt Gödel**

---

Mathematicians, logicians, and philosophers of the twentieth century looking to demystify the mind proposed that all human thought and reasoning were the products of purely mechanistic or algorithmic processes. This model of human cognition envisioned the brain as a machine performing computation, and in that sense, it was an intellectual advancement, but it suffered from being overly reductionistic.

In 1936, the inventor of the digital computer, Alan Turing, showed that a relatively simple machine capable of performing mechanical operations could compute the solution to any problem that could in principle be solved with computation, given enough time and memory. This hypothetical machine would come to be known as a Turing machine. Our computers and smartphones are for all practical purposes Turing machines.

If the brain were simply an organic Turing machine, then everything that the mind does—all the mental magic associated with conscious thought—could be reduced to logical operations (rule-based symbol manipulation). The mind would be no more mysterious than a digital computer, which we understand completely. It was a pretty straightforward way to look at brain and mind, but the brilliant Austrian logician Kurt Gödel would prove that view wrong, in a quite roundabout and loopy way. To understand how he did it, we are brought back to epistemology, the study of how we acquire true knowledge about the world.

Before Gödel's famous incompleteness theorem was published, mathematicians and logicians assumed that all statements about the world could either be true or false, and in theory there should be an algorithmic way of determining which category a particular statement belonged to. According to this philosophy, mathematical proof was the source of true knowledge. This black-and-white conception of reality is a product of the nature of their work. A mathematical conjecture, like the famous theorems we hear about (e.g., Fermat's last theorem) can either be true or false, depending on whether or not a mathematical proof exists for that particular proposition.

It may be ridiculously difficult to find that proof, but in theory, if a certain mathematical statement is true, it should have a logical proof, and if it is false, then no proof should exist. At least that was the assumption made by leading mathematicians like David Hilbert, Bertrand Russell, and Alfred North Whitehead, who tried their very best to design the ultimate formal system that could quickly prove or disprove any conceivable mathematical conjecture. Despite a valiant attempt to achieve this goal, Gödel demonstrated that it was an impossible task, calling into question the power of mathematics to explain all of reality, as well as the assumption that the mind is simply a machine.

## The Unprovable Paradox

With a clever trick, Gödel cast doubt on the assumption that truth and mathematical proof are equivalent. He took an ancient Greek paradox involving self-reference, known as the *liar's paradox*, and converted it into

a mathematical statement about number theory using an absurdly complex coding scheme, known as *Gödel numbering*. Although the theorem is extremely complicated, we can comprehend it by translating it into equivalent statements in normal language. To understand how the liar's paradox inspired a new kind of proof—one that would seem to call into question the foundations of mathematics—you really have to explore it for yourself. Consider the following sentence, which is a less-confusing variant of the statement "I am lying," and try to determine whether it is true or false:

"This statement is false."

Go ahead and say the statement out loud and see how trying to prove it true or false sends you in a loop that does neither. If it is true, then the statement must be false because it says that it is. But if the statement is false, then of course it cannot be true. What happens when we assume the statement is false? Because the statement itself states that it is false, if the whole statement is false, it would mean that the statement must be true. But it can't be true if it says it is false!

Either road inevitably leads to a contradiction, but it is only by going around the loop that you see the nature of the paradox. Because the statement *refers to itself*, we get an unprovable statement that is neither true nor false. It cannot be proved because proving it would disprove it. Now go ahead and try the same with the sentence "I am lying."

You may have noticed that this is really just a language game—the paradox only occurs because the sentence has the absurd property of talking about itself. That may be so, but it doesn't change the fact that self-referential mathematical statements expose a vulnerability in any logical system claiming to be the determiner of all truths. No formal system can be considered *consistent and complete* if it produces "undecidable" conjectures —mathematical propositions that cannot be proved true or false. Hence the name *incompleteness theorem*.

But the real genius of Gödel's theorem was not that it created a mathematical statement that was neither true nor false, like the self-referential statements "This sentence is false," or "I am lying."

Gödel bumped the loopiness up a level by constructing a mathematical statement that was *true but unprovable*. Of course, to mathematicians of the time, this would sound like a flawed claim, as in their minds truth could

only be determined by provability. So, how did he do it? Consider the following self-referential statement, which is not just about itself, but its own provability:

"This statement has no proof."

Here's the thing about this particular proposition: It is in fact true; it has no proof. And we know that it can't have a valid proof without even checking or giving it any thought, because if a mathematical proof did exist, it would prove the statement true, but the statement says it has no proof; so again, the statement can't be proved true without disproving itself. Despite not being provable, we can clearly see that it's true. It is a truth that doesn't need an actual proof because its truth is self-evident. It states that it has no proof, and it does not.

Okay, so that's kind of quirky and neat, but what does this all have to do with the mystery of mind? It turns out that Gödel's incompleteness theorem was not only about mathematics but also the mental processes of mathematicians.

## The Mind Is Not a Turing Machine

Gödel's true but unprovable proposition suggests that there are mathematical truths that exist outside the realm of what can be determined using symbolic logic or computed by a Turing machine. A mathematician could temporarily get around Gödelian incompleteness by extending the fundamental axioms of the formal system to include a class of self-referential statements. But Gödel clearly showed that such an effort would be futile, because that new iteration of the system would contain new kinds of true but unprovable self-referential statements. You can never plug the leak that is self-reference, unless you are willing to keep extending your mathematical model forever.

With this in mind, we can see that Gödel's great theorem was hinting at some sort of theory of everything like the unifying theory of reality described by poetic meta-naturalism, which is forever incomplete only because *reality itself is open*. That is to say, nature is a continual work in progress, thanks to emergence. Emergent phenomena, and the mathematics used to describe them, such as causal calculus and information geometry,

must be added on to any fundamental physics theory if that theory claims to be a true theory of everything. And even then we will not be done. More exotic mathematics will continually get added to the model of reality as new kinds of dynamical laws and complex behaviors emerge. For a formal system or a scientific theory to capture more truth about reality, for it to acquire more knowledge in an effort to completely describe nature, it must continuously adapt, just like a biological system.

The whole program ignited by Gödel was a big blow to reductionism, because reductionism is based on the idea that there is a single mathematical theory that can capture everything about the universe in a perfectly precise and complete way. That dream must be abandoned. Gödel's theorem also dealt a big blow to materialism by showing that the brain is not simply a Turing machine.

In his 1989 classic book *The Emperor's New Mind*, the decorated theoretical physicist Sir Roger Penrose revived an argument put forth by the philosopher John Lucas three decades earlier. Because human mathematicians can see the truth of an undecidable statement that is the numerical equivalent to "This statement has no proof," the mind seems to be doing something beyond raw computation:

> The inescapable conclusion (of Gödel's theorem) seems to be: Mathematicians are not using a knowably sound calculation procedure in order to ascertain mathematical truth. We deduce that mathematical understanding—the means whereby mathematicians arrive at their conclusions with respect to mathematical truth—cannot be reduced to blind calculation!

While Penrose and Lucas are often credited for this keen insight, it is clear that Gödel himself was well-aware of this implication, having said many times a version of the following quote: "Either mathematics is too big for the human mind or the human mind is more than a machine." What's the difference between mind and machine? Machines compute; minds *understand*. They can see the truth in an unprovable statement, a truth not accessible to a purely algorithmic intelligence. What enables this curious ability we call understanding? Conscious experience, presumably.

While Penrose correctly interpreted this to mean that the mind is not simply a Turing machine (a digital computer), he made the arguably unjustified leap to assuming that the brain must then be a quantum computer, or at least some kind of organ that exploited the quantum mechanical property known as superposition to solve problems and make decisions. While this theory, developed in collaboration with the anesthesiologist Stuart Hameroff, should not be dismissed on the grounds that it invokes a quantum explanation, at the present moment it is not taken seriously by most consciousness researchers.

The main argument against the hypothesis, put forth by physicist and integrated-information-theory proponent Max Tegmark, is that any coherent quantum state would collapse or "decohere" much too quickly to influence any neural processes needed for consciousness because of how warm, wet, and noisy the brain is. However, recent findings show that numerous biological processes critically involve quantum processes—like photosynthesis, which uses quantum tunneling, and bird navigation, which makes use of quantum entanglement. If quantum biology is now universally accepted as real, quantum neurobiology could be just around the corner. A general rule about evolution is that if there's a computational advantage to some phenomenon in nature, natural selection will usually find a way to leverage it.

While Gödel's incompleteness theorem made consciousness more mysterious to Penrose, self-reference was the key to solving the puzzle for Douglas Hofstadter.

## Self-Reference as the Source of Self

Hofstadter made Gödel part of pop culture with his Pulitzer Prize–winning book *Gödel, Escher, Bach: An Eternal Golden Braid*, which was released in 1979, a decade before Penrose's book. More than forty years later, it still provides fresh insights to philosophers of mind, artificial intelligence researchers, and neuroscientists looking to understand the computational correlates of consciousness. What Hofstadter saw in Gödel's theorem was that self-reference created paradox because the symbols in a Gödel sentence take on *meaning* when they are *about themselves*. Suddenly, there is more

than just syntax; there is also semantics. In the preface to *Gödel, Escher, Bach's* twentieth-anniversary edition, Hofstadter wrote:

> Something very strange emerges from the Gödelian loop, the revelation of the causal power of meaning in a rule-bound but meaning-free universe. And this is where my analogy to brains and selves comes back in, suggesting that the twisted loop of selfhood trapped inside an inanimate bulb called a "brain" also has causal power—or, put another way, that a mere pattern called "I" can shove around inanimate particles in the brain no less than inanimate particles in the brain can shove around patterns.

So, Hofstadter believes that the self, the I, the conscious observer, has true causal power on the world. He also gives us a vague but powerful insight into how this observer capable of causation comes into existence. By his account, the self emerges from self-reference "via a kind of vortex whereby patterns in a brain mirror the brain's mirroring of the world, and eventually mirror themselves, whereupon the vortex of 'I' becomes a real, causal entity."

Hofstadter often uses an interesting metaphor to explain how self-reference in the form of self-modeling produces something like a soul.

> For an imperfect but vivid concrete analogue to this curious abstract phenomenon, think of what happens when a TV camera is pointed at a TV screen so as to display the screen on itself (and that screen on itself, etc.)—what in GEB I called a "self-engulfing television," and in my later writings I sometimes call a "level-crossing feedback loop." When and only when such a loop arises in a brain or in any other substrate, is a person—a unique "I"—brought into being. Moreover, the more self-referentially rich such a loop is, the more conscious is the self to which it gives rise.

This loopy idea, that consciousness emerges from self-modeling, has found support from some influential figures in the machine-learning world. Judea Pearl, whose causal calculus is providing the math needed to describe high-

level causation, echoed Hofstadter in a 2019 interview with MIT podcast host Lex Fridman.

> That's consciousness. You have a model of yourself. Where do you get this model? You look at yourself as if you are a part of the environment . . . I have a blueprint of myself, so at that level of a blueprint I can modify things. I can look at myself in the mirror and say, "Hmm, if I tweak this model I'm going to perform differently." That is what we mean by free will.

Fridman probes the question further, asking, "What do you think consciousness is? Is it simply self-awareness—including yourself into the model of the world?"

"That's right," Pearl confirms. "Some people tell me, 'No, this is only part of consciousness,' and then they start telling me what they *really mean* by consciousness, and I lose them. For me, consciousness is having a blueprint of your software."

But talk is cheap. What does the neuroscience say?

## The Neural Correlates of Consciousness

To understand how feedback loops and self-modeling create consciousness and instill systems with genuine free will, we must first understand how the brain is organized. For this, it is undeniably useful to think of the brain as a type of computer, but it is very different from the computers we are familiar with (which have a "von Neumann architecture"). Its parallel design means that the brain's computing is distributed across many modules. There are about one hundred billion neurons in the brain, and each neuron is connected to about ten thousand other neurons. In the prior chapter, we talked about hierarchical modular architectures in biology. Cells are integrated complexes of many interacting biomolecules, multicellular organisms are complexes of cells, and societies are complexes of organisms. The brain, as a complex adaptive system built by evolutionary transitions, is also organized hierarchically, both structurally (in terms of anatomy) and functionally (in terms of processing).

In this hierarchical architecture, neurons are the fundamental processing unit, and together they carry out collective computation. Neurons that commonly fire together are organized into groups called *neuronal assemblies*, which form larger brain networks that have specialized information-processing functions. For example, there is a region in the brain for processing faces called the *fusiform face area*, and a region that interprets speech, called *Wernicke's area*. These areas consist of relatively high-level processing because they get information that is sent up through the processing hierarchy from brain regions that decode more basic features of a stimulus, like form and color. Remember, not all brain states are accompanied by conscious experience. Consciousness dissolves when we go to sleep and remains absent until we enter a dream. So what is it that makes conscious states computationally special?

Details aside, the brain generates conscious experience when coordinated global activity emerges from the local electrical interactions of billions of neurons—the synchronous firing of those neurons integrates information from multiple processing streams into a single field of experience. It is not that synchronous firing of any kind produces conscious states, as we also see long-range synchronies with epilepsy. So, it is more careful to say that coordinated global firing at a particular synchrony (thought to be ~40 Hz) is *necessary but not sufficient* for consciousness. This global activity is made possible by a phenomenon we are now quite familiar with, thanks to cybernetics: feedback. Feedback loops running from the thalamus to the cortex—so-called thalamo-cortical loops—integrate information and bind features into a cohesive perceptual landscape.

When does consciousness disappear? When the feedback loops disappear, because it is these loops that entrain the activity of the component parts, bringing about the global integration of information. Whenever there is feedback, there is some form of self-reference at play, and in brains, it can be a signature of self-modeling. Without feedback, the brain still functions as a physiological organ controlling autonomic functions, but consciousness fades into nonexistence. We know the mental model is still encoded in the neuronal architecture of the brain, but what disappears is the observer, the self, which comes from the self-reference we get from real-time self-modeling. When the strange loop is disrupted—the

multilevel feedback loop that mirrors itself—consciousness fades. So, the mind should no longer be thought of as a nonphysical entity moving things around; it is a multilevel cybernetic controller with causal power over itself and its environment.

This description is still vague, but even the strongest skeptic would surely agree that we are making progress on the hard problem of consciousness, which is how a material brain creates subjective experience. What do modern neuroscience theories of consciousness add to this picture, if anything?

## Feedback Loops Create Consciousness

According to integrated information theory, a purely feed-forward system, which is a system that has no feedback loops that integrate information, can in theory mimic any conscious behavior without having the corresponding conscious experience that a system integrating information would have. It is important to note this does not necessarily mean such a system would exhibit the same type of flexible intelligent behavior that a conscious system would; it simply means that any particular response associated with conscious thought can be duplicated without consciousness. This suggests loops are a necessary ingredient for consciousness in both organisms *and* machines.

For example, the cerebellum is a brain region that is complex and sophisticated, yet it does not seem to produce a conscious observer, despite being the structure that houses the majority of the brain's neurons. Anything you do that requires coordinated movement involves the cerebellum, like typing, walking, dancing, or driving. These activities don't typically require conscious control or effort, and some of them, like typing, are performed much worse if you try to think about what you're doing. If the cerebellum is clearly processing information, then why isn't there "something it feels like" to be a cerebellum? Christof Koch explains why in a 2018 *Nature* article titled "What Is Consciousness?":

> The cerebellum is almost exclusively a feed-forward circuit: one set of neurons feeds the next, which in turn influences a third set. There

are no complex feedback loops that reverberate with electrical activity passing back and forth.

I am reminded of a startling moment from my childhood when I saw my sleeping grandmother's hand shoot up toward the ceiling and remain there as she continued snoring. I later learned that it was not the first time that happened, and although it is an unusual example, it nicely demonstrates how an action that is normally attributed to conscious or voluntary effort can be programmed and executed without consciousness. And since it happened routinely, I think it is safe to say that the motor movement was not the result of her consciously raising her arm in a reoccurring dream, although it is impossible to say for sure.

Larissa Albantakis, an articulate defender of integrated information theory, published in 2014 and 2015 a series of simulation studies that suggest you could have humanlike robots that seem conscious in every way but really have no mental experience at all. One system would have a feed-forward design (involving no feedback loops) and the other would integrate information through feedback. Both systems would execute the same kind of behavior when presented with the same sensory stimulus, though only the latter would be having a subjective experience. Both machines would tell you they were conscious, but one would be lying. It sounds like sci-fi, but math doesn't lie. According to this theory, conscious machines would require what is called *neuromorphic hardware*, which duplicates the brain's causal dynamics at the physical level, rather than merely simulating the brain digitally on a standard von Neumann architecture, which would not produce consciousness for the same reason computer simulations of black holes do not actually bend space-time around your computer.

## Global Neuronal Workspace Theory

Around the same time that Tononi was developing his theory of information integration, another quantitative theory of consciousness was gaining popularity—cognitive neuroscientist Bernard Baars's global workspace theory, which would be updated by Stanislas Dehaene some years later to produce global neuronal workspace theory, a more neuroscientifically

accurate version of the computer-inspired cognitive architecture. Global neuronal workspace theory proposes a computational model for how separate processing streams are integrated into a unified conscious perception through emergent synchronous activity that produces a "global workspace." If the global workspace is the mental workspace that we call the conscious mind, how exactly is that workspace created?

We can answer that question by recognizing that the workspace corresponds to a neural attractor, and loops are obviously involved in the formation and maintenance of any attractor. It is specifically fronto-parietal loops that create the global workspace—or loops between functionally connected regions in the frontal lobes and the parietal lobes. When the activity of neurons in these areas become coordinated past a certain critical threshold, waves of global order propagate through the network, synchronous brain-wide firing occurs, and a global attractor of persistent electrical activity emerges from a phase transition that Dehaene calls an "ignition event."

This phase transition in the brain broadcasts the information in the global workspace to other neural modules that typically process information separately, leading to what Daniel Dennett calls "fame in the brain." Now the information in the conscious representation is available for all kinds of additional processing not possible without the global workspace. Specifically, information in the workspace becomes the focus of *selective attention*, and its contents can be manipulated by working memory, or it can be used to make high-level decisions that the brain has no stored solution for. Planning and imagination are also made possible by the global workspace, which momentarily subsumes those independently working systems into the neural attractor, allowing minds to do things that can't be done without consciousness, like running mental simulations of possible futures in order to select the one that is most desirable. Why should we think the global workspace has something to do with self-modeling? Because, again, we see feedback loops, and when the loops disappear, so does the mental workspace. Conscious processing of globally available information is disrupted and higher-level cognition dissipates.

So, the neuroscience suggests that Hofstadter is right. It is the causal loops running across levels that create consciousness, what neuroscientist Anil Seth characterizes as a "controlled hallucination," which is a step up

from “illusion,” as Dennett described it in his landmark 1991 book *Consciousness Explained*. Some neuroscientists and philosophers have jokingly called the book “Consciousness Ignored,” because it explained many things about how the brain processes information without saying much at all about how subjective experience is created. Do not be fooled by the language games philosophers play to perpetuate controversy—the mental simulation you are experiencing right now is as real as anything else in reality, because conscious experience has causal power.

The mind can move things around in a paradox-free manner because it is not an immaterial entity as once thought; it is a cybernetic control unit that can steer itself toward goals that keep it in existence. Dennett himself now argues that cybernetic control is the source of free will, although he prefers to describe control without talking about consciousness. The problem with trying to leave consciousness out of the discussion is that it obscures the fact that there are different mechanisms of control, with different neural and computational correlates, which have different degrees of causal power. Conscious control is a more sophisticated mode of self-regulation that emerges later in evolution, when a cognitive metasystem transition creates a high-level controller with the powers of the prefrontal cortex.

## Integrated World Modeling Theory

This sounds like a satisfying account of the neural underpinnings of consciousness, but there is just one problem. Proponents of integrated information theory, like Koch, have identified specific areas in the back of the brain, the so-called posterior hot zone, as being the neural correlates of conscious experience, while the global neuronal workspace is associated more with the front of the brain. Only one theory can be right. Right?

A high-profile program designed to test these two seemingly incompatible theories is currently underway, but there might not actually be any conflict between the two models at all. In 2020, a young neuroscientist at the Johns Hopkins Center for Psychedelic and Consciousness Research named Adam Safron introduced an ambitious theory of consciousness called *integrated world modeling theory*, which says that in addition to

integrating information, the emergence of an observer requires that a system be able to generate a model of self and world with "spatial, temporal, and causal coherence." In this model, integrated information, like synchronous neuron firing, would be necessary but not sufficient for consciousness, doing away with integrated information theory's panpsychism problem, described in Chapter Ten.

Safron's model attempts to unite integrated information theory and global neuronal workspace theory with Friston's free-energy principle and the associated active inference framework, which we learned about in previous chapters. The conflict between Tononi and Dehaene's theories can be resolved by recognizing a distinction first made by the philosopher Ned Block in 1995. According to Block, these neuroscientists are talking past each other because they are talking about two different things. In Block's words, *phenomenal consciousness* refers to "experience; the phenomenally conscious aspect of a state is what it is like to be that state."[1] This kind of consciousness is associated with feeling, emotion, and perception, while *access consciousness* refers to higher cognitive function, like focused attention on a task, idea, or problem. When you completely space out to music, only phenomenal consciousness is involved, but if you start thinking about what the song means, access consciousness is activated.

We also know that humans can override their normal subjective experience with imagination, as we do when we are asked to describe how a friend looked the last time we saw them. In this case, it would seem that the contents in access consciousness override the contents of phenomenal consciousness.

The posterior hot zone described by Koch would be the neural correlates of phenomenal consciousness, while the prefrontal areas described by global neuronal workspace theory would be the neural correlates of access consciousness. These two networks would be responsible for distinct types of conscious processing, but they would interface with each other to coordinate conscious cognition and rational behavior. If this story is correct, Hofstadter's strange-loop model of consciousness would be substantiated. A level-crossing feedback loop would connect access consciousness to phenomenal consciousness (cortico-cortical loops) and both to the thalamus (thalamo-cortical loops), which is

the brain's relay center and the decider of where in the cortex sensory information gets sent to be processed. The strange-loop concept can explain how predictions made by higher brain regions, like the prefrontal cortex, can influence the streams of sensory information coming in from the thalamus. In a 2021 interview, Safron explained the connection between consciousness and Bayesian inference.

> With respect to the Bayesian brain, in brief the idea is that perception is a kind of inference to the best guess as to the causes of sensory observations, so that your perceptual experience is a kind of prediction or probabilistic inference of what you think is in the world. How is this related to consciousness? The ignition events that create global workspaces represent a selection of the winning interpretation or model of what's happening in the world, relative to a range of alternatives.[2]

Conscious observers require feedback loops because they integrate information, but this is not enough for consciousness. That is to say, like synchronous firing, integrated information is necessary but not sufficient. The mind must have mental contents, must be about something, and must have a world model built up through evolution or adaptive learning. For AI systems, this means being embodied, preferably as humanoid robots. What the loops do is conjure up a mental simulation of that world model, allowing the agent to run simulations of possible futures, so that the agent can select the path that it predicts is most consistent with its long-term goals. The global workspace, then, would be associated with a higher level of control, which grants new freedoms to the system, such that it can truly choose its fate. That ability is what we are talking about when we talk about free will.

## Clearing the Confusion About Free Will

If free will is real but restricted, just how free are we? The famous "free will experiments," conducted in the early eighties by American neuroscientist Benjamin Libet, suggest that some of the decisions we think

we make consciously have already been decided for us. In these studies, participants were asked to voluntarily choose when to make a movement and to note the precise moment in time they feel the choice was made. Monitoring the brain activity associated with this response revealed that the feeling of making a voluntary decision is always preceded by a spike in brain activity, called the *readiness potential*, that can be used to predict when the movement would be made. So, it would seem that before we make a decision, the brain has already decided for us! This suggests that most of the time we don't actually have conscious control over our behavior—we are like viewers in a theater, along for the ride but unable to steer the ship.

These results have been commonly misinterpreted to mean that neuroscience has debunked free will. The first thing to say is that even if the decision to move is made by the unconscious mind, the movement can be halted or altered by the conscious mind, which exerts control over the preprogrammed behavior. In a 1999 article titled "Do We Have Free Will?," published in the *Journal of Consciousness Studies*, Libet wrote: "The volitional process is therefore initiated unconsciously. But the conscious function could still control the outcome; it can veto the act. Free will is therefore not excluded. These findings put constraints on views of how free will may operate; it would not initiate a voluntary act but it could control performance of the act."

These findings are consistent with our model of the mind as a multilevel controller. If you are only able to do what is programmed by the older and more primitive networks in the brain, that would not seem very free, because you'd be like the bacterium obligatorily performing chemotaxis. A causal agent becomes "free" from its preprogrammed agenda when a higher level of cybernetic control emerges in its cognitive machinery that allows the system to analyze its own actions.

Conscious causation would refer to the ability to override our automatic behaviors, which are all those programmed in genetically, as well as those that have been learned through lived experience and encoded as procedural memory. Most of our actions and decisions don't require any conscious deliberation, so it is really the movement driven by contemplation and reflection that we are talking about when we speak of free will.

In fact, a 2019 study conducted by Christof Koch and colleagues showed that when the experiments designed by Libet involve decisions of

importance, rather than arbitrary decisions—like when to make a pointless movement—the readiness potential disappears. While arbitrary decisions are aimless, unreasoned, and without consequence, deliberate decisions would have significant practical or moral consequences.[3] The abstract states:

> We directly compared deliberate and arbitrary decision-making during a $1000-donation task to non-profit organizations. While we found the expected RPs [readiness potentials] for arbitrary decisions, they were strikingly absent for deliberate ones. Our results and drift-diffusion model are congruent with the RP representing accumulation of noisy, random fluctuations that drive arbitrary—but not deliberate—decisions. They further point to different neural mechanisms underlying deliberate and arbitrary decisions, challenging the generalizability of studies that argue for no causal role for consciousness in decision-making to real-life decisions.

In a 2018 interview with MIT AI researcher and podcast host Lex Fridman, Koch, a neuroscientist trained as a physicist, made it very clear that he believes that a choice based on conscious deliberation is an exercise of free will:

> There are hundreds of experiments that show how we can be influenced. But in the final analysis, when you make a life decision —I'm talking about critical decisions—when you really think: should I marry, should I go to this school or that school, should I take this job or that job, should I cheat on my taxes or not. These are things where you really deliberate and I think under those conditions, you are as free as you can be. When you bring your entire conscious being to that question and try to analyze it under all the various conditions, then you make a decision, you are as free as you can ever be.

So, it seems that although being conscious does not necessarily mean an agent has free will, consciousness is what enables free will. An agent is free

when it has the ability to consciously and deliberately control its future actions. More specifically, free will is when an agent can exercise control over a preprogrammed behavior that the high-level controller—the conscious mind that monitors an organism's actions—determines to be maladaptive or suboptimal behavior. Suboptimal behavior would be any behavior that is deemed inconsistent with the agent's long-term goals. A couple examples can clarify the difference between behavior that is instinctual or reflexive and behavior that is regulated or controlled.

If you are walking through a market and are starving because you skipped lunch, you might want to grab that delicious-smelling kabob off a truck and sink your teeth in, but as a civilized man or woman, you don't do that unless you have the money to pay for it. But if we brought a caveman to Times Square in a time machine, he would probably have a very hard time keeping his hands off the food, and if he's in a "red light" district, off of other things. To be civilized means to transcend our primitive brain and reflexive behavior, which is not typically going to produce the optimal response for a member of modern society. This requires an additional level of control, which is enabled by the global workspace of the prefrontal cortex.

When you think about the spectrum of different types of human behavior, we see that free will is something we possess only if we have healthy executive functioning and frontal lobe activity, a fact that is of the utmost relevance to the justice system. But if that activity becomes impaired, the high-level controller vanishes and free will goes along with it. When this happens, an agent literally has less or no control over their patterns of behavior, and we see this clearly with addiction. Addicts who cannot override their tendency to do something that is hurting them are essentially stuck in an unhealthy behavioral attractor that they cannot break free from. The neuroimaging literature shows that cocaine addiction, for example, impairs networks in the front of the brain associated with cognitive control and self-monitoring. As a result, addicts are "less free" and more of a slave to their limbic system. On the brighter side, studies suggest that meditation and cognitive-training exercises can enhance activity in frontal networks associated with self-control, which appears to be something like a muscle that can be strengthened.

We also see people lose their free will in mental disorders where frontal lobe activity is impaired. There are reports from schizophrenics that describe feeling like they are being propelled by forces beyond their control. This presumably arises when the neural mechanisms underlying cognitive control dissolve, leaving us with old brain structures that are executing wired-in programs. There is no high-level controller monitoring action, so you will again get behavior that is not in the best interest of the agent and not consistent with its long-term goals. Brain-imaging studies show that schizophrenia is associated with changes in prefrontal circuitry that negatively affect higher cognitive functions like working memory and attention.

I believe some similar impairment can explain the experiences of people with the curious mental disorder known as Cotard's syndrome. Those unfortunate souls experiencing Cotard's delusion believe they are dead, and in many cases, die of starvation because "the dead do not have to eat." Sure enough, brain imaging studies show impaired frontal lobe activity, specifically in regions associated with cognitive control, in such patients. Perhaps the feeling of being dead arises when one feels they have lost agency or causal power, leading them to conclude they are ghosts.

## The Curious Compatibility of Destiny and Free Will

Now we see why consciousness requires more than just integrated information. A sentient system must have a cohesive model of the world, built up through interaction with it, and that world model must include a model of the system doing the modeling. The loveably loopy gospel of multiple realizability tells us there is more to intelligence than the size of an agent's mental repertoire, as measured by Tononi's phi. Levels of awareness seem to be crucial for high-level intelligence and cosmic-scale causal power. If this is true, then AI researchers should pay more attention to the hierarchical modular design of the brain and the self-modeling it performs to produce reflective thought, such as that which can make sense of true but unprovable Gödel sentences. Doing so may lead to conscious machines that

think, feel, and love. Whether such machines would need a biological core is an open question.

Simon DeDeo, a complexity scientist at Carnegie Mellon University and the Santa Fe Institute, has a message for those working in the field of machine learning: "Great progress in physics came from taking relativity seriously. We ought to expect something similar here: success in the project of general artificial intelligence may require we take seriously the relativity implied by self-reference."[4]

So, what progress have we made in our understanding of the mind-body problem? The solution to this philosophical puzzle comes from recognizing that information is physical, despite being naively categorized as nonphysical in the past, which created an unresolvable paradox. Information is always encoded in a material substrate and cannot exist apart from that substrate, and that includes the information being manipulated by human minds. This should be no more mysterious than the fact that software programs cannot exist in any meaningful sense without hardware to implement them. That being said, in the same way that computers running software produce digital simulations that have their own reality in a virtual space, brains running "mindware" produce mental simulations that have their own intrinsic reality. Minds can't be reduced to brains, but they also cannot be separated from them. It seems more appropriate to say that minds are an emergent property of systems that are configured to process and integrate information.

When it comes to nature, there is simply a mechanical level, a computational level, and with certain types of computation, an experiential or phenomenological level, all corresponding to one physical reality, and all having an equally justified *ontological status*. In other words, they all represent properties that are equally real. To be clear, all experiential (psychological) processes are also computational and mechanical processes, but not all mechanical or computational processes produce experience. Mind is not a new form of matter; it is a phase of ordinary matter that is unlocked when biological machinery discovers a special type of configuration, specifically one that encodes a model of itself. Don't let the various levels of description confuse you, though. These coexisting properties all emanate from a single natural process. Nature simply has

levels that emerge through evolution, and in more ways than one: physical levels that create an observable hierarchy of nested cybernetic systems, culminating at the planetary scale in a biosphere equipped with a global brain, and more abstract levels that represent properties that cannot be captured with mathematical description alone. Experience requires qualitative description in addition to computational processes. Words are symbols that map to phenomenological phenomena—the things we experience but can't reduce to numbers without losing the essence of the thing we set out to describe—like sensations, perceptions, and emotions.

The story does not stop with the level of agency associated with self-awareness and free will, now understood to be conscious cognitive control. A higher level of self-modeling, also associated with the prefrontal cortex, allows humans to simulate themselves in a seemingly infinite variety of scenarios. With a sufficiently sophisticated global workspace comes the power of *imagination*, which enables us to play with the mental model of the world that evolution and adaptive learning have built up in our brains. We can travel back and forward in time in our mind, and this allows us to see beyond the present and into the future. It is our ability to see the existential challenges we face decades down the line that motivates our civilization to invent the technology needed to avert otherwise inevitable disaster—the technology that spreads adaptive complexity to other planets.

In this way, the continuation of cosmic evolution toward higher complexity and greater knowledge crucially depends on intelligent agents that can and will exercise free will. It is the power to simulate counterfactuals, or "what if" scenarios, that will illuminate the causal path that is consistent with the continued persistence of life in the cosmos. By using our conscious mind to imagine the future implied by an open-ended evolutionary process, we can see that intelligence-aided recursive self-organization will continually add new levels to the growing hierarchy of life and mind, making metasystems of increasing computational complexity, causal power, and self-referential richness. That gives us a duty to fulfill as sentient citizens of the cosmos.

In the model of reality proposed by poetic meta-naturalism, there is room for destiny *and* free will. Because the future is not determined in the strict sense imagined by Laplace, there is no paradox. As individuals, we are truly free to choose our future path, but because it is in our nature to try

and survive, our actions are not random; they are directed toward a goal, a goal that we all share as living agents. This shared goal—evading equilibrium—harmonizes our collective activity, and because human civilization learns from its mistakes, progress becomes inevitable.

Since cosmic evolution inexorably generates increasingly robust and powerful multilevel cybernetic control systems, the emergence and outward expansion of intelligence in the cosmos was *predestined*, but in a completely natural sense of the word, as Christian de Duve, Freeman Dyson, Harold Morowitz, Ray Kurzweil, Christof Koch, and other renaissance men of science have argued. The freedom is at the level of the agent who exercises conscious causal control, not the aggregate behavior of humanity, which settles into increasingly stable and integrated attractors as knowledge accumulates in cultural memory.

Our unifying theory of reality has sent us around some loops, but in doing so, we have made progress in understanding life's origin, evolution, and destiny. Now we are ready to ask what comes next in the grand and majestic process of cosmic self-organization.

# 13

# Transcendence and Enlightenment

---

> I assume that something new will emerge in human society, and it will present us with undreamed possibilities in science and the arts. This emergence requires our efforts and requires something spiritual that goes beyond the mind. There will be a new emergence, and we will play a part in what that emergence is. That is our destiny.
>
> **—Harold Morowitz, *The Emergence of Everything***

---

We are nearing the end of our journey, so it is a good time to quickly go over what we have learned. The integrated evolutionary synthesis, which serves as our unifying theory of reality, is based on the evolutionary epistemology-universal Darwinism-universal Bayesianism framework, which was developed in Part Two of the book in detail. This entirely-too-long name just clarifies that all three paradigms are complementary perspectives corresponding to the same cosmic story.

Universal Darwinism emphasizes that the universe is evolving at every scale through both competitive evolution and through self-organization, which is cooperative evolution among agents that have formed a collective

unit. Evolutionary epistemology emphasizes that the evolutionary process is a problem-solving procedure that creates knowledge. Universal Bayesianism is a Bayesian update of the first two paradigms. It emphasizes that the knowledge encoded in biological systems is stored in a world model. This model is an internal representation and statistical mapping of the environment, which gets updated through adaptation and adaptive learning.

We have explained how conscious minds with the capacity to experience and feel the world emerge from collections of neurons processing and integrating information. A unified field of subjective experience requires that an adaptive system have a sufficiently rich internal model of the world, which must include not only the biologically relevant variables in the surroundings but also the agent doing the modeling. The "self" that is the observer arises from the self-referential act of self-modeling. By introducing the idea of an internally embedded world model, the unifying theory of reality builds a bridge between matter, mind, and cosmos. Hierarchical emergence through recursive self-organization inexorably creates complexity and knowledge that produces an observer, an observer who emerges specifically when the system starts to model itself.

We have explained the mechanisms underlying agency and free will, and how complex adaptive systems can impact the physical world around them through intentional and purposeful action, all in perfect accordance with natural law and the logic of causality. And we have learned that considering the phenomenon of consciousness in the context of cosmic evolution can help explain its mysterious nature. Mind is not a quirk; it is the modality through which the cosmos wakes up. Mind is how matter gets to experience existence.

Last but not least, we have poetic meta-naturalism, a natural philosophy of reality, based on the gospel of multiple realizability, that says that the universe organizes itself through loops and levels. Multiple realizability means the meaning is to be found in the metaphor, because the metaphor reveals the substrate-independent pattern being realized at different scales. Since complex adaptive systems at different scales have similar dynamics, and evolve through similar developmental cycles, knowledge of one level provides a model for phenomena at other levels. For example, we get a better understanding of what a society or "social organism" is in terms of its

dynamics and how it evolves because we see how it is like a biological organism. In the same way, we get a better understanding of an ecosystem by recognizing how it is like a chemical autocatalytic set. There is some truth to the statement "There is nothing new under the sun," because each new level that emerges is a "rhyme" on the level before it. But because recursive self-organization and multilevel emergence are always bringing novel patterns and properties into the world, we find that "history doesn't repeat itself, but it rhymes." This makes the future semi-predictable. We cannot predict the precise behavior of any single individual, but we can develop what Isaac Asimov called a "social statistical mechanics" to understand how humans behave in aggregate.

As our cosmic journey continues, the story is going to speed up, because it is the story of cosmic evolution, and the rate of evolution is always accelerating. What does the paradigm predict for the future of humanity and the future of the cosmos as a whole?

## The Global Mind Wakes Up

One important thing to understand, which was explained in the prior chapter, is that studying a complex adaptive system at one level can lead to insights that help us better understand functionally equivalent systems at other levels.

Now that we've understood mind a little more, let's return to the topic of the global brain. If we are to take the idea of multiple realizability (substrate independence) seriously, then minds can be realized on different mediums and on different scales, as long as the relevant causal and computational mechanisms are duplicated in brain-like hardware. If this is true, we must entertain the idea that an information-processing system like the internet, which in this definition includes the humans it connects, can possess conscious states similar to those generated by a biological nervous system. As long as information is being integrated by feedback loops that generate models of the world and the self—in this case, the "self" being the system that is the global brain—mind will presumably emerge.

At the current time, it seems highly unlikely that the internet possesses first-person experience. Our brains have been shaped by evolution over

millions of years in ways that have refined and expanded our information-processing capabilities and world model. While the global brain may not be self-aware, or even aware, there's certainly a collective intelligence at work—humanity has coordinated its behavior enough to warrant that description. I suppose the real question is: Could *someday* the collective human intelligence "wake up," producing a coherent planetary mind? And, if so, would our individual consciousness disappear as we slip into the collective consciousness, or would there be conscious entities nested within conscious entities? If so, how would control be distributed exactly? These are romantic questions, indeed.

Integrated information theory's current iteration says that for a global mind to emerge from the human collective, the amount of information being integrated by the internet (measured by phi) would have to exceed the amount of information being integrated by the individual processing units—the human brains that act as neurons in the global brain. But when that event happens, and there's no reason to think it won't eventually, our minds should merge into one mysterious and majestic übermind, but only for as long as we stay physically connected to the whole. This possibility has been explored by Christof Koch:

> Consider humankind's largest and most complex artifact, the Internet. It consists of billions of computers linked together using optical fibers and copper cables that rapidly instantiate specific connections using ultrafast communication protocols. Each of these processors in turn is made out of a few billion transistors. Taken as a whole, the Internet has perhaps $10^{19}$ transistors, about the number of synapses in the brains of 10,000 people. Thus, its sheer number of components exceeds that of any one human brain. Whether or not the Internet today feels like something to itself is completely speculative. Still, it is certainly conceivable.[1]

What does our unifying theory of reality, in the context of poetic naturalism, say about this question? It says we cannot know for sure *what* will emerge at these higher levels, but we can be fairly certain that *something new* will emerge. What comes after the mental? If the

mechanistic mystic Pierre Teilhard de Chardin was right, the emergence of the noosphere, or the global mind, will elevate the mental experience of humans to something that can only be described as spiritual.

It is important to keep in mind that the global mind, should we bring it about by coordinating our activity in the correct way, is not the end goal of the evolutionary process; it is merely a step on the ladder of open-ended complexity growth. The major evolutionary transitions of biology are just a subset of a larger chain of cosmic metasystem transitions, which are far-from-equilibrium phase transitions that continuously create new levels of hierarchical control. Darwinian dynamics are at work above the level of biological evolution, propelling cultural and technological evolution forward, making a self-replicating biosphere as inevitable as was the origin of life, given enough time and enough evolutionary cycles. When intelligent life terraforms a new planet, that is the biosphere replicating, and because the planet will have different properties than the planet of origin, there will be replication with variation.

## Loops + Levels = Lift Off

Through loops and levels, continual scientific and technological progress will inexorably lead to the outward expansion of life into the cosmos. Barring any species-annihilating catastrophes, our knowledge and technological prowess will eventually allow the human race to transcend our biological limitations. Post-biological beings that are capable of recursively improving their design—making every iteration more intelligent and resilient than the previous one—will result in what Ray Kurzweil calls an "intelligence explosion" in his book *The Singularity Is Near*. This meta-evolution, an evolution of evolutionary mechanisms, will make spreading through the galaxy and beyond attainable for our unfathomably tech-savvy descendants, as well as any post-biological civilizations that have emerged on other planets. Maybe rather than "post-biological," I should call these future agents "hyperbiological," because transcending human limitations does not mean dumping biological design, which we know has been optimized in many ways by natural selection.

According to Kurzweil, not only would a sufficiently advanced civilization rapidly proliferate and expand through space, as a matter of necessity, they would also transform the matter and energy in their midst into forms that allow them to meet the metabolic and computational demands required for further expansion. Integrating "dumb" matter into their computational structure will suffuse it with patterns of information, and with them, intelligence and sentience.

Because complexity and knowledge both grow at an increasing rate, continual technological progress will enable life and consciousness to spread through the cosmos at accelerating speeds. As long as life can continue to find new sources of free energy, this process can continue. If our unifying theory of reality is correct, doing so will not only be hard, it will take every last ounce of effort for life to overcome these existential challenges. It will always seem like we are on the brink of disaster, but because *problems create progress*, cosmic evolution proceeds and adaptive complexity spreads.

For example, the fact that our sun will explode in a few billion years creates a looming deadline for intelligent life on Earth. The second law of thermodynamics and stellar lifetimes provide something like a game clock for the game of life. To persist, life must progress at an increasingly fast rate, but it will do so because it is constantly adapting and modeling the world around it with increasing detail and accuracy. Because knowledge accumulates, life learns to control the cosmos for its own benefit.

In a paper published in 1960, the distinguished physicist Freeman Dyson popularized a sci-fi concept that came to be known as a *Dyson sphere*, a structure built around a star that would enable an intelligent civilization to meet their energy requirements after exhausting resources at home. With billions to trillions of stars in each galaxy, and over a hundred billion galaxies in the observable universe, there are plenty of solar batteries in the sky for intelligent life to feast on. Since it'll be at least five billion years until the sun burns out, humans have more than enough time to evolve into hyperintelligent post-biological beings capable of improving themselves recursively, transcending mortality and all its constraints. This new cosmic narrative forces us to conclude that life is not merely insignificant "chemical scum," as Stephen Hawking disparagingly put it in *A Brief History of Time*. Through Darwinian evolution and the knowledge

creation that comes from it, life becomes increasingly proficient at manipulating matter and energy in ways that will ultimately allow it to influence the unfolding of cosmic events.

If this is indeed the inevitable fate of the cosmos and destiny for sentient life, the question is how long can it go on for? What is the ultimate fate of life in the universe? Now we will explore the limits of computation and ask the question, "Can information processing and consciousness go on forever?"

## A Happy Ending Means No Ending at All

While it is commonly presumed that all life in the cosmos will end in an unavoidable "heat death" or "big freeze" inevitably brought on by an ever-expanding, increasingly entropic universe, some of our most respected scientific minds do not believe that is necessarily the case. Various cosmic models—put forth by esteemed theoretical physicists including David Deutsch, Seth Lloyd, Lee Smolin, Paul Davies, Freeman Dyson, Julian Barbour, George Ellis, John Barrow, and Katherine Freese—describe how the fate of sentient life could turn out to be more fortunate than what is commonly assumed. Can a self-aware system, with an information-processing capacity that approaches the infinite, engineer an endless state of subjective experience that stretches on forever? Any answer to this question will admittedly be speculative, as all our theories of the ultimate fate of the universe are a work in progress. However, these hypotheses are scientifically rigorous and our best guesses, given the evidence.

Lloyd, author of *Programming the Universe*, believes that increasing complexity is built into the laws of nature, and although this statement seems harmless, it has the grandest of implications. Since Einstein proved matter and energy are interchangeable, Lloyd explains in his book how intelligent life could collect all the matter attainable in an expanding universe to maintain its existence:

> Trillions of years from now, the stars will have burned through their store of nuclear fuel. At that point, our descendants, should they still be around, could harvest energy by collecting matter and converting

> it into usable energy, a strategy analyzed in detail by Steven Frautschi of Caltech. The maximum amount of free energy that could be extracted is $E = mc^2$, where $m$ is the mass of the collected matter.

Lloyd's book discusses multiple scenarios in which life in the cosmos could have a happy ending, which would mean no ending at all:

> By scavenging farther and farther afield, our descendants will collect more and more matter and extract its energy. Some fraction of this energy will inevitability be wasted or lost in transmission. Some cosmological models allow the continued collection of energy ad infinitum, but others do not . . . A more parsimonious strategy for eternal life is to make do with a finite amount of energy, as proposed by Freeman Dyson, of the Institute for Advanced Study.

In a landmark article for the journal *Reviews of Modern Physics*, Dyson showed how intelligent life in an expanding universe could perform an infinite amount of computational operations given a finite amount of energy, and physicists David Deutsch and Frank Tipler proved the same for a closed model of the universe. This means that even in a universe that eventually ends in heat death, or a Big Crunch—the former appearing much more likely—an infinite *subjective* existence for life is still possible if certain conditions are met.

The observation that the universe is expanding at an accelerating rate would seem to make meeting these requirements a Sisyphean task, as shown by physicist Lawrence Krauss. But a paper in the journal *Physics Letters B* by astrophysicists Katherine Freese and William Kinney, titled "The Ultimate Fate of Life in an Accelerating Universe," demonstrated that if the rate of expansion begins to slow, the amount of energy within the horizon will continue to increase indefinitely. Needless to say, an endless supply of free energy is good news for life.

Some physicists have argued that dark energy is the game changer. In his highly praised book *The Beginning of Infinity*, Deutsch wrote:

> Depending on what dark energy turns out to be, it may well be possible to harness it in the distant future, to provide energy for knowledge-creation to continue forever. Because this energy would have to be collected over ever greater distances, the computation would have to become ever slower . . . the inhabitants of the universe would notice no slowdown, because, again, they would be instantiated as computer programs whose total number of steps would be unbounded.

Deutsch is very clear about the potential open-ended nature of evolution: "This process need never come to an end. There are no inherent limits to the growth of knowledge and progress."

Kauffman made a similar argument in his 2016 book *Humanity in a Creative Universe*:

> The second law says free energy is running down. But we know now that the expansion of the universe is accelerating due to the mysterious dark energy that comprises about 70 percent of the energy of the entire universe. The implications of this accelerating expansion is that we do not have to worry about enough free energy. As the universe becomes larger, its maximum entropy increases faster than the loss of free energy by the second law, so there is always more than enough free energy to do work.

## Reaching a Maximal State of Complexity

Even more fantastical than life simply continuing forever, might there be some kind of grand culmination to universal self-organization? Some have theorized that cosmic evolution—which would seem to roughly mirror a biological developmental process—must inevitably lead to the emergence of a maximally complex, maximally integrated, maximally aware state of the cosmos. In other words, a great cosmic mind at the end of time, so computationally powerful and so capable of generating new realities and experiences that it is functionally indistinguishable from the notion of a god. It would be the most powerful god conceivable. An ultimate *Omega*

*Point*, to use Teilhard de Chardin's term for a system that has reached the goal state of maximal complexity.

Physicist Paul Davies wrote: "Many scientists have speculated that, as the time-line stretches towards infinity, so an emerging super-intelligence will become more and more god-like, so that in the final stage, the super-mind will merge with the universe: mind and cosmos will be one."

He went on to say, "Then the whole character of the universe, including the emergence of its laws and the nature of its states become inextricably intertwined with its mentality—with its mindfulness."[2]

In this grand vision, it is not just that life can potentially go on forever; in its struggle to do so, it inevitably evolves into a god, and if we call it anything else, we are playing semantic games.

While the idea of an emerging superintelligence that stretches out into infinity is fascinating and full of philosophical implications, it isn't worth contemplating if it isn't possible. Is such a state both attainable *and* sustainable?

According to physicist Sean Carroll, a trusted name among atheists and skeptics, there is no fundamental reason to believe cosmic evolution has an end. "If there exists a maximal entropy (thermal equilibrium) state, and the universe is eternal," wrote Carroll on his popular blog *Preposterous Universe*, "it's hard to see why we aren't in such an equilibrium state—and that would be static, not constantly evolving. This is why I personally believe that there is no such equilibrium state, and that the universe evolves because it can always evolve."

Distinguished cosmologist Lee Smolin suspects that the universe itself is a self-organizing system. In his 1997 book *The Life of the Cosmos*, he said:

> What we know about cosmology allows us to consider the universe as a whole to be a self-organized system . . . It is clear that if the natural state of matter is chaos, an external intelligence is needed to explain the order and beauty of the world. But if life, order, and structure are the natural state of the cosmos itself, then our existence, indeed our spirit, might finally be comprehended as

created naturally, by the world, rather than unnaturally and in opposition to it.

At this point, we can only speculate about which ultimate scenario is correct—or most likely. But we can say for sure that the cosmos is gradually waking up through life, it does so inexorably, and it is not clear whether this process ever has to end.

## A Natural Worldview for the Spiritual and the Secular

A universe that is naturally evolving toward an increasingly complex and conscious state illustrates that, contrary to past scientific consensus, we do in fact live in a universe that progresses toward a very particular kind of state that can be appropriately called the "goal" of the system. While there may be no final goal, but instead an unending series of ever-grander attractors, there is still a predictable trajectory toward higher complexity. This does not mean the universe had a conscious intent prior to the emergence of intelligent agents, because it did not—agency, intent, and purpose emerged through evolution. But they did so inevitably, as a consequence of the "cosmic code."

The universe has a goal in much the same way an oak seed does, as it will inevitably develop into a tree, or an embryo into an adult. Of course, these analogies aren't perfect, but the fundamental laws and constants of physics play the functional role of a genetic program that guides development. A goal-directed universe, such as a universe that grows increasingly intelligent and self-aware, is described in philosophical terms as being teleological. The cosmos certainly has a *telos*, and if we ignore that, we will be missing out on a paradigm with unrivaled predictive and explanatory power. It allowed Teilhard de Chardin to predict the emergence of the internet before the computer had even been invented.

Cosmic teleology is a worldview that recognizes the goal-directed nature of our universe, a dynamical system that is becoming increasingly complex and sentient as time moves forward and natural processes play out.

The cosmic teleology presented in this book is an entirely *natural teleology*—specifically an evolutionary-computational teleology that is fully compatible with the laws of physics—and a scientifically rigorous theory, as it can be described in quantitative terms and makes testable predictions.

Because in the past the word "teleology" has sometimes been associated with a religious belief in a supernatural or mystical force that guides evolution toward progress, the label will undoubtedly bother some skeptics and atheists who have an aversion to religion. But this should not deter us from using the most appropriate and recognizable terminology. It is worth reiterating that the cosmic self-organization process being described here is entirely mechanistic—a sort of "blind groping" toward an increasingly complex state, driven by the universe's proclivity to process information and accumulate knowledge in the architecture of adaptive complexity. It is not guided by any external force or metaphysical entity; functional organization is internally generated from the physical laws and the evolutionary dynamics that emerge from the constraints those laws impose on matter in motion.

American philosopher Thomas Nagel explicitly called for a natural teleology in his controversial 2012 book *Mind and Cosmos*, and he should be commended for his bravery. But Nagel made the crucial mistake of suggesting that we need "nonphysical laws" to explain this aspect of our world and was rebuked for it by a number of prominent scientists and book reviewers. Nagel made his argument vulnerable by using a paradoxical term: If there are such nonphysical laws in existence, upon recognizing such laws, they would just get naturalized and added to the list of phenomena we consider real and "physical." For example, information was generally described as nonphysical until we finally understood what information was and how it was encoded. Then we saw that it, too, was physical. So rather than classifying these emergent "laws of life" as nonphysical, it is more appropriate to describe them as *new laws of physics*, as Schrödinger suggested in his book *What Is Life?* These are the laws and principles from non-equilibrium thermodynamics, cybernetics, evolutionary theory, computational theory, and dynamical systems theory.

Poetic meta-naturalism addresses this issue, because it recognizes that what is considered "natural" will change over time, as new phenomena emerge, like life, mind, and technology—and as we formalize these

phenomena with mathematics and computational models. Because our models of reality will always have some uncertainty and incompleteness, the line between the physical and metaphysical will never be completely clear. Today's metaphysics might be tomorrow's physics. For example, by traditional scientific standards, the concept of a multiverse should be considered a metaphysical idea because it is an untestable theory, at least at this moment in time.

This violates Popper's requirement that a scientific theory be falsifiable. Still, many of today's leading physicists believe that the multiverse exists, despite no empirical evidence for it. An exception is the radical reductionist Sabine Hossenfelder, who has gone as far as to compare belief in a multiverse to belief in a religion. We will come back to this topic shortly. The point is that as we understand more about the world, and start to formalize complex phenomena that once seemed beyond science, the realm of the "physical" and "natural" will expand, and it will do so forever, due to the emergence of novel phenomena with new kinds of properties. It may also expand due to us finding out that reality is larger and more complex than we initially assumed.

A modern cosmic teleology—a computational or Darwinian teleology—eliminates all "explanatory gaps" in the evolutionary story, shattering the creationist's claim of "irreducible complexity," and in turn putting the nail in the coffin for any "intelligent design" theories that invoke the supernatural to explain life's origin. But reducing the teleological development of the cosmos to a series of natural mechanisms does not explain away the mystery of why we find ourselves in a self-organizing world that becomes increasingly self-aware and computationally powerful through a series of inevitable emergences. We no longer have "irreducible complexity" in the creationist's sense, but we have *irreducible emergent phenomena* like life and consciousness, which can't be explained away reductively—that is, in terms of atoms following fixed trajectories—without losing something critical in the explanation.

So now we have explained the inevitable emergence of life and progressive evolution toward higher complexity scientifically without any reference to anything supernatural, but we are left with something like a cosmic program that we must account for. In this case, "unweaving the rainbow" only makes it more mysterious—and more romantic. As the new

cosmic narrative resolves age-old existential conundrums, such as how we got here and where we're going, it brings with it new ones that will require much thought and discussion.

For example, if our universe turns out to be just one of many, will these other universes have a tendency toward increased complexity and knowledge, or is ours unique? Does a self-organizing universe imply that we are living in a simulation, or something analogous to it? If so, does this suggest that there was an intentional agent, or collective of agents, that designed the hardware and software of the system—in other words, set the laws, constants, and initial conditions—and then let the system evolve according to its own dynamics?

The line between simulation theory and a mechanistic "intelligent design" theory (life is generated from internal laws that appear fine-tuned) is subtle if not nonexistent. If we can create simulated agents in simulated worlds, who is to say that we aren't such a creation? Might the end state of the universe that Paul Davies described serve some sort of function in a more comprehensive "base reality"? While such speculations are still squarely in the realm of philosophy, it is clear that science is revealing a universe that is rational, comprehensible, and self-actualizing.

Let's explore this idea a little further, since for many, it is the most important question of the new cosmic narrative. Why does the universe appear to be fine-tuned for the emergence of intelligent life?

## The Fine-Tuning Problem

To understand why an inherently metaphysical theory like a multiverse theory was embraced by mainstream physics, we need a final brief history lesson. The discovery of the apparent "fine-tuning" of the physical constants uncovered by theoretical physicists in the 1970s created a problem for atheism that some scientists desperately wanted to solve. Cosmologists, such as Sir Martin Rees and Paul Davies, explained that if the value of a certain parameter of our universe, like the strength of the gravitational force, or the electromagnetic force, or the speed of light, were altered only slightly, we would not just have a universe with no life, we'd likely have a universe with no planets or stars or physical objects either.

Stephen Hawking wrote: "The laws of science, as we know them at present, contain many fundamental numbers, like the size of the electric charge of the electron and the ratio of the masses of the proton and the electron . . . The remarkable fact is that the values of these numbers seem to have been very finely adjusted to make possible the development of life."[3]

Daniel Dennett, one of the most widely respected atheists and skeptics, described the fine-tuning mystery this way in *Darwin's Dangerous Idea*:

> As more and more has been learned about the development of the universe since the Big Bang, about the conditions that permitted the formation of galaxies and stars and the heavy elements from which planets can be formed, physicists and cosmologists have been more and more struck by the exquisite sensitivity of the laws of nature. The speed of light is approximately 186,000 miles per second. What if it were only 185,000 miles per second, or 187,000 miles per second? Would that change much of anything? What if the force of gravity were 1 percent more or less than it is? The fundamental constants of physics—the speed of light, the constant of gravitational attraction, the weak and strong forces of subatomic interaction, Planck's constant—have values that of course permit the actual development of the universe as we know it to have happened. But it turns out that if in imagination we change any of these values by just the tiniest amount, we thereby posit a universe in which none of this could have happened, and indeed which apparently nothing life-like could ever have emerged: no planets, no atmospheres, no solids at all, no elements except hydrogen and helium, or maybe not even that—just some boring plasma of hot, undifferentiated stuff, or an equally boring nothingness. So isn't it a wonderful fact that the laws are *just right* for us to exist? Indeed, one might want to add, we almost didn't make it! Is this wonderful fact something that needs an explanation, and, if so, what kind of explanation might it receive?

To the dismay of many atheists, this reopened the debate about a cosmic designer. But this would be the god of *deism*, which is specifically not a creator that intervenes in the physical world, and that includes the world of human affairs. This stance is not to be confused with the "intelligent design

movement," which argues that specific events in the history of the universe, such as the origin of life, required a supernatural act by God. This is an unfortunate label that has led to much confusion because an entirely physical theory involving an intelligent creator, designer, or programmer that doesn't intervene is consistent with science.

This model would include theories that propose that we are living in a simulated universe, since they imply some kind of intelligent programmer. Such theories are modern-day versions of the deistic view of many giants of math and science, including Isaac Newton, Gottfried Leibniz, James Clerk Maxwell, Arthur Eddington, Max Planck, Werner Heisenberg, and Kurt Gödel, just to name a few. The deity of these men of science was *God the architect of the physical laws and constants*, inventor of the mechanistic universe, which evolves and unfolds as natural processes play out. Famed atheist Richard Dawkins happily admits that this concept of God is a reasonable one, because God is simply a sufficiently advanced intelligence. If we can create simulated worlds, there's no reason to believe it hasn't been done before. Dawkins said, "We could take a deist god, the sort of god of the physicist—the god of somebody like Paul Davies—who devised the laws of physics; God the mathematician, God who put together the cosmos in the first place, and then sat back and watched everything happen. The deist god—I think one could make a reasonably respectable case for that. Not a case that I would accept, but I think it is a serious discussion that we could have."[4]

In the same interview, Dawkins went on to say: "You could possibly persuade me that there was some kind of creative force in the universe. That there was some kind of physical, mathematical genius, who created everything; the expanding universe, devised quantum theory, relativity all that. You could possibly persuade me of that."

So, is the finding that the laws and constants are in some sense fine-tuned for the development of life evidence for an intelligent designer? Not necessarily. There is an explanation that has kept most atheists happy, but we will see that this explanation comes with a lot of philosophical baggage.

## Fine-Tuning Solutions

The fine-tuning problem is explained in two different ways with two different versions of the *anthropic principle*, which correspond to opposing physical theories, both of which would be deemed inherently metaphysical theories a century ago because neither is directly testable, at least at the current moment in time. Again, this violates Popper's requirement that a scientific theory be falsifiable. However, this is no excuse to ignore the possible explanations to the fine-tuning mystery, because mysteries justify theories that could provide explanations that solve the mystery. Remember, today's metaphysics could be tomorrow's physics.

The *strong anthropic principle* says that the parameters of the universe are finely tuned so that it can produce life and mind. In this teleological model of reality, the universe evolves according to a program that ensures that life will spread through the cosmos, either forever or until something like a cosmic Omega Point emerges. According to the strong version of the anthropic principle, nature has an intentional design and is intrinsically purposeful and meaningful. While this view might sound mystical, the ascent of life in the universe, and the tendency for the cosmos to self-organize, are computational processes that inevitably occur due to a "cosmic code" that is analogous to a genetic code, but this developmental plan is encoded in the laws and constants of physics.

If our universe is the only universe in reality, it would be extremely improbable to find ourselves in a fine-tuned cosmos by chance. The strong anthropic principle implies design, and it therefore attributes the fine-tuning of the laws and constants to some kind of intelligent creator or programmer. Sean Carroll, who does not support the strong anthropic principle or believe that our universe was created by an intelligence who intended to produce a cosmos with life, nevertheless states that the fine-tuning argument is "the best argument we have for God's existence." In *The Big Picture*, Carroll openly admits that "the existence of life provides, at best, a small boost to the probability that theism is true." The same could be said for deism, which differs from theism in that if there is a creator, it is one that never intervenes, so the laws of physics are never violated. Carroll sees the doctrines of theism and naturalism as fundamentally opposed, though deism and naturalism need not be.

Whether the creator imagined by deism is a god or a technologically advanced being (or collective intelligence) would seem to be a purely

semantic issue, as Freeman Dyson pointed out: “I do not make any clear distinction between mind and God. God is what mind becomes when it is passed beyond the scale of our comprehension.”[5]

Again, there is no fundamental difference between the idea that our universe is a simulation created by hyperadvanced intelligent beings and the idea that it was created by a god with unimaginable powers. Of course, if the universe was designed by an intelligent agent, this begs the question: Who or what created that agent? This appears to lead to an infinite regress that scientists and philosophers do not like. Is the reality that God inhabits also fine-tuned for the inevitable emergence of God? If so, we are faced with the same problem we had before invoking God. And, if we could somehow manage to explain how God emerged through logical processes, then we could just use that same mechanism to explain our own emergence and existence, again making God an unnecessary part of the explanation. Since this explanation for the fine-tuning mystery has some major problems, let’s move on to the alternative, the *weak anthropic principle*.

The weak anthropic principle on its own doesn’t explicitly take a stance on the issue, but it does so implicitly. It is a tautology that simply says we find ourselves in a universe that appears finely tuned for life, because if the fundamental constants weren’t conducive to life, no conscious agents would have emerged who could ponder such questions. If there are observers who can ask questions about their origin, those observers will always find themselves in a world that has laws and constants that have values that would permit their existence. In other words, the fine-tuning of our universe’s parameters should not be surprising—it is the only way things could be, given the fact that we are here!

## The Multiverse Conundrum

So, does the weak anthropic principle tell us anything interesting? Well, if you take the position that there’s only one universe and this is all there is to reality, then you have the burden of explaining why, out of all possible universes, the one that exists happens to be one that supports life and consciousness and, if you accept the thesis of this book, the open-ended

growth of complexity and the continual creation of knowledge. But if you don't assume that our universe is the only one in existence, the fine-tuning problem seems to disappear. The idea of a *multiverse*—a vast collection of a very large if not infinite number of universes—provided a logical solution for the scientists and philosophers who opposed the religious-seeming implications of the strong anthropic principle.

According to the multiverse explanation, we find ourselves in a universe that appears to be intentionally fine-tuned for life, but that is only an illusion, because our universe is just one of many in an overwhelmingly lifeless multiverse. In this model, there is a universe out there for every possible combination of values of physical constants, and purely by chance, a few produce something interesting, like life. This position, popular with atheists, reduces life's significance once again and restores the idea that reality in its totality is purposeless and mostly, if not entirely, meaningless. Conscious agents who can ponder existential questions are mere blips in the vastness of space and time, anomalies destined for transience.

Today, many leading theoretical physicists and cosmologists support the idea of a cosmological multiverse, despite an inability to test the theory empirically. Since universes are *causally isolated*—meaning they do not interact in any way—it is impossible to directly measure anything that is outside our universe. How interesting: An inherently *metaphysical* theory is accepted as a *physical* theory! Why? Because it seems to be the only good explanation for the fine-tuning mystery that does not entail an intelligent creator.

This reveals a profound insight that has been mostly overlooked by philosophers. The lesson to be learned is that at any particular moment in time there is no clear line between the physical and metaphysical. For the same reason that a metaphysical theory like a multiverse theory can become science, a theory that posits a creator, what one might call a supernatural theory, can also become a natural theory, if we consider the possibility that this creator inhabits its own physical reality, out of which it emerged by mechanistic processes. We will come back to this possibility shortly.

While the weak anthropic principle's multiverse explanation for the fine-tuning problem satisfies some, like MIT's Max Tegmark, many scientists and philosophers have pointed out that this explanation has its own baggage, with many of its implications being harder to swallow than

the idea that our universe was designed by an intelligence. According to Tegmark's model, in the vastness of the multiverse, despite being mostly lifeless, there exists every logically possible universe, which includes all those that you can imagine as well as all those you can't. This includes universes in which you pursue a career as a clown, and universes where, just by chance, a statistician trying to understand the laws of probability flips a coin a thousand times and they all come up heads. You can imagine how distorted the science might be in a universe that just happens to feature massive chains of unlikely events.

In Tegmark's multiverse, there are also universes where intelligent beings emerge through self-organization and Darwinian evolution and go on to create simulated worlds with simulated agents that are functionally indistinguishable from "real-world" agents. In other words, the multiverse is so vast, that even though the phenomenon of life is relatively rare, there are still so many life-producing universes that there are world creators, or world programmers, in some of those universes. Thus, the *cosmological multiverse* (as opposed to the "many worlds" multiverse, which will be discussed later) includes universes that have intelligent designers and universes that produce intelligent designers. So, for some observers at least, the simulation theory or the "God theory"—which we see now are essentially the same thing—would be true, but it would be *trivially true* rather than *the ultimate truth*.

Essentially, the multiverse theory associated with the weak anthropic principle says "everything imaginable must exist somewhere," and it is hard to see how this is a superior solution to the fine-tuning problem than that proposed by the strong anthropic principle. In fact, the strong anthropic principle narrative can logically exist inside the multiverse model, which would seem to be problematic, though one could interpret this to mean that the two versions of reality are logically compatible.

## Cosmological Natural Selection

Now that we've considered both of the major explanations for the fine-tuning mystery, which one is most compatible with the new cosmic narrative that has been articulated in this book? What theory best explains a

self-organizing universe where life not only emerges but spreads throughout the cosmos such that the entire universe can be seen as waking up through the conscious agents its laws and parameters have manifested?

A statistical argument against the multiverse theory might go as follows: If the multiverse explanation favored by atheists is correct, and the universe is mostly lifeless, then there should be a relatively small set of universes that are friendly for life—though the size of this set could be staggeringly huge compared to the numbers we are used to thinking about. Now, out of all of the life-friendly universes, there will presumably be universes where only unicellular life emerges, and universes where intelligent life emerges but fails to ever leave the planet, as well as universes where intelligent life continuously evolves and expands forever. Paul Davies calls universes where life is permitted but is transient and insignificant *minimally biophilic* universes, and universes where life influences cosmic evolution *optimally biophilic* universes.

Minimally biophilic worlds are merely life friendly while optimally biophilic worlds are teleological—constantly evolving toward greater intelligence. This distinction is absolutely critical, because if we are in fact in an optimally biophilic universe, as the new cosmic narrative presented in this book suggests, that would present a big problem for the multiverse explanation and the weak version of the anthropic principle. Why does the multiverse solution to the fine-tuning problem fail, given that we can see that the evolutionary process in our universe is open-ended?

Intelligent universes will presumably require greater fine-tuning, so we'd expect the set of minimally biophilic universes to be vastly larger than the set of optimally biophilic universes. In other words, there will be many more universes where life emerges but is transient and cosmically insignificant. Now here's the crucial point. If there are a far greater number of merely life-friendly universes compared with full-blown teleological universes, then statistically we would expect to find ourselves in a universe where life is transient and cosmically insignificant. The fact that we inhabit a teleological universe where knowledge creation can (and presumably will) continue forever would seem to favor some flavor of the strong anthropic principle. "Luck" just doesn't cut it.

For those not ready to accept the quasi-religious implications of the strong anthropic principle, there is a third alternative that would explain

why teleological universes are so common that we should not be surprised to find ourselves in one. This third option also happens to be completely in line with the Darwinian nature of reality that has been the unifying theme of this book. Furthermore, it is an explanation that our must trusted atheists and skeptics find attractive, like Richard Dawkins and Daniel Dennett. This third way is provided by a theory called *cosmological natural selection*, and it may be the best explanation we have for our fine-tuning mystery, a mystery that becomes more mysterious when one realizes that life is central to the universe's large-scale evolution.

Thanks to Stephen Hawking's work on *singularities*, we now know that black holes and the Big Bang have a common physical attribute. They are both points where the curvature of space-time approaches the infinite. Our universe appears to have emerged from a singularity (the Big Bang) and then proceeded to manufacture singularities (black hole formation). Lee Smolin's big idea is that when a black hole forms in our universe, it creates a "baby universe" on the other side, through a new big bang event. Smolin speculates that the newly created universe would inherit the physical parameters (laws and constants) from its parent universe. However, since nature is thermodynamically noisy at a fundamental level, there would presumably be something analogous to a mutation that causes a small tweak or variation to the constants or laws of the baby universe, so that it is either more or less fine-tuned for the growth of complexity. Daniel Dennett describes it this way:

> The basic idea is that the singularities known as black holes are in effect the birthplaces of offspring universes, in which the fundamental physical constants would differ slightly, in random ways, from the physical constants in the parent universe. So, according to Smolin's hypothesis, we have both differential reproduction and mutation, the two essential features of any Darwinian selection algorithm. Those universes that just happened to have physical constants that encouraged the development of black holes would *ipso facto* have more offspring, which would have more offspring, and so forth . . .[6]

If one is to take this idea seriously, as respected cosmologists like Leonard Susskind have called for, then we are left with a new kind of multiverse: One that starts out with a single universe but constantly grows as offspring universes are created and differentially reproduce. It is a *Darwinian multiverse* that evolves and generates emergences. If black-hole formation is the method of reproduction, then natural selection would favor universes that were the best at creating black holes. What's most interesting about this theory is that because the conditions conducive to black-hole formation are the same conditions that are friendly for carbon-based life, for technical reasons that Smolin explains in *The Life of the Cosmos*, this cosmological reproduction process inevitably produces biophilic universes where life emerges as a robust phenomenon.

If cosmological natural selection is followed to its logical conclusions, it does not just favor life-producing universes but also universes with intelligent life. Specifically, technologically savvy intelligences who know how to engineer black holes, which is done in the service of perpetuating life. This possibility has been explored by physicists like Louis Crane and Paul Sorensen, as well as complexity theorist James Gardner in his wonderful 2003 book *Biocosm*. In *Physics Today*, Smolin wrote:

> Paul Sorensen proposes an application of cosmological natural selection (CNS) according to which intelligent beings would be motivated to create artificial black holes, which would tune the ensemble of universes to favor those that are hospitable to intelligent life. Versions of that idea were proposed by Louis Crane and Edward Harrison, although Sorensen's version is different in proposing that it could explain why the universe's laws are comprehensible to us. His idea is elegant, but let's wait till the few predictions of CNS are thoroughly tested and confirmed . . . before engaging in speculation about implications that will be hard to test.[7]

Some of the superstars of cosmology, including Alan Guth and Martin Rees, have speculated that it would be quite possible to manufacture black holes or baby universes using technology like particle accelerators in the far future. If this method of creation is possible, then there's no reason to believe that intelligent life won't someday achieve it. And if we achieve it,

and demonstrate that such a thing is possible, it would increase the likelihood that our universe is also the result of intelligent engineering.

Cosmological natural selection is consistent with the unifying theory of reality and makes universal Darwinism beyond universal. It is attractive to evolutionary theorists because it explains how a complex universe that is finely tuned for life emerges from a reality that starts off simple: a lifeless universe that produces a black hole. While some atheists may embrace this theory as a refutation of the strong anthropic principle, upon reflection, it also describes a reality that is teleological, or life generating and purposeful. In the evolving multiverse described by the cosmological natural selection theory, life not only emerges inevitably but also quickly comes to dominate the multiverse such that the growing ensemble of universes are mostly if not entirely optimally biophilic universes that further perpetuate life and mind. So, this model of reality ultimately has the same teleological structure as our universe.

It seems that our best explanations for the fine-tuning mystery describe a reality where life, intelligence, and consciousness have fundamental significance and grow without limit. Now we can ask the question: Is there a god? Yes, there's an evolutionary process that will inevitably and continually create minds of increasing power that are indistinguishable from gods. Are we the products of such a god? We may be, and we may become one. But all gods emerge as a result of an open-ended evolutionary process that creates functional forms of increasing computational sophistication. So, it seems more accurate to view the process *itself* as God. From the view of poetic meta-naturalism, God is an algorithm for eternal recursive emergence.

## Making Sense of Quantum Mechanics

It is now common knowledge that at the quantum scale—the microscopic scale that is impossible to see with the naked eye—reality gets weird. Particles not being observed or measured seem to exist in multiple places at once, a phenomenon known as *quantum superposition*. This means that at the quantum scale, an electron (for example) has no concrete existence and no well-defined location in space. It is as if the physical objects at this level

are smeared out into a blurry cloud of potential existence, making it unclear whether the word "physical" is still an appropriate adjective. This cloud of possibility is described by a mathematical structure called a *wave function*.

But when we try to measure where that electron is, suddenly the wave function "collapses," and the previously smeared-out electron takes on a tangible existence in a precise point in space. The wave function represents a statistical distribution that describes the probability that one possible position will manifest relative to the other possibilities when the quantum system is probed by a scientist's measuring device. This seems to suggest that reality is not entirely deterministic but has a random, probabilistic, or stochastic nature to its evolution. Upon the act of measurement, the system becomes part of the classical world—the concrete world we experience—which somehow emerges from the weird quantum world of superposition.

The nature of this wave function collapse is still mysterious to physicists because we don't completely understand why simply observing a quantum system forces nature to choose a definite outcome and why that specific outcome occurs instead of other possible outcomes. This mystery is known as the *measurement problem*, and the founders of quantum mechanics, Einstein among them, considered the problem to be one of the most significant questions regarding the ultimate nature of reality.

If the quantum-to-classical transition only occurs when a quantum system is being observed or measured, does this mean that physical reality is not actually "out there" when we aren't looking at it or probing it in some way? Einstein famously said, "I like to think that the moon is there even if I am not looking at it." After all, like every other physical object, the moon is made up of matter consisting at bottom of subatomic particles, which are quantum systems. Why isn't the moon in a state of blurry superposition when it is not being observed?

Einstein's other quantum quip was "God does not play dice," and it concerned the other aspect of the measurement problem. If the observed outcome of the measurement process is truly random, it would seem to suggest that the evolutionary trajectory of the classical world we experience is at least partially determined by chance events. The future would not be precisely determined, and although this seems to be good news for ideas like agency and free will, on its own it implies that what actually happens in the future is merely a result of the roll of the dice.

Einstein had a big problem with this idea of chance in nature because it contradicted his religion—the religion of pantheism, which sees God in the inherent harmony and unfolding order of the natural world, a world that he believed was both logical and comprehensible. A number of "hidden-variable theories" have been proposed that attempt to show that wave function collapses are not truly random events, and that the universe evolves in a completely deterministic way. If we could find such a theory, we would know why a given quantum superposition collapses into one state rather than another upon measurement. However, all the proposed hidden-variable theories have been refuted by experimental findings, all which seem to suggest that wave function collapse is a truly probabilistic event.

So, what are the potential solutions to the measurement problem that modern science has to offer? Today's most popular solutions are the many-worlds interpretation of quantum mechanics and a model of wave function collapse known as *decoherence*. Though these models are not incompatible, they each can exist without each other, so they will be explained as independent ideas.

The many-worlds interpretation is one that a growing number of leading physicists find attractive because it does away with the measurement problem completely. This interpretation says there is no wave function collapse that eliminates possibilities in favor of one definite outcome. Instead, at the moment a measurement occurs and we perceive a definite outcome in our reality, the universe branches into a number of parallel universes, with each universe representing one of the possible outcomes specified by the wave function. So, if you are a scientist making a measurement on a quantum system, after the measurement there will be one universe where you observed one possible outcome, and another where you have observed another, and another where you have observed another. In other words, all of the possible ways the evolution of the universe could go are realized in some branch of the many-worlds multiverse. In this picture, the entire multiverse follows a deterministic trajectory where branching is constantly occurring, and the only thing that is random is which universe you happen to find yourself in after the measurement. It's certainly a weird idea, but what about quantum mechanics is not weird?

There are a number of problems with this explanation, though. For one, like the cosmological multiverse theory, these supposed other universes are

not directly measurable, making the model impossible to test scientifically, at least at the current time. The other problem is that it sounds absolutely bonkers: there is now a "you" that exists in a growing number of parallel universes. This again leads to the idea that all logically possible universes exist, including universes where you became a professional clown as well as universes where you solved the measurement problem.

## Refuting the Many-Worlds Hypothesis

The same statistical argument that was used to refute the cosmological multiverse hypothesis can be used to refute the many-worlds hypothesis. If the many-worlds interpretation is correct, we wouldn't expect to find ourselves in a universe where intelligence evolves and expands through space inevitably and open-endedly. There are many logical scenarios where intelligent life on Earth keeps failing to get off the planet for one reason or another. But if the new cosmic narrative presented in this book is correct, and adaptive complexity (life) can't be killed or contained because it is constantly learning and adapting, then the evolutionary trajectory of our universe is not one that is arbitrary or random. As a result of recursive emergence, the cosmos is continually self-organizing and moving in the direction of increasingly complexity and computational power.

Finding ourselves in a universe with a teleological trajectory (an optimally biophilic universe) when most conceivable paths through the many-worlds multiverse are those where life is transient (minimally biophilic universes) would again seem highly improbable. So, either the many-worlds multiverse does not exist, or all or most of the parallel worlds that harbor life are also teleological and moving toward ever-greater complexity and intelligence. We cannot rule out the latter speculation, since it is possible that any world that abides by the second law of thermodynamics also generates increasing complexity in non-equilibrium conditions. If you find this whole line of reasoning to be far-fetched, you are not alone.

Fortunately, there is another alternative explanation that addresses both of Einstein's quantum mechanical qualms, and it involves the process of *decoherence* mentioned earlier. Decoherence explains why the moon is

really there whether we are looking at it or not, and why external reality exists independent of observers who construct measuring devices to study natural phenomena. The macroscopic world we observe around us does not contain large-scale objects in a superposition of states—unlike a subatomic particle, a chair is either here or there, but not both—because quantum-wave functions are constantly being collapsed by their physical environments, which could be thought of as nature's measuring devices. For example, any given atom at any given time is in contact with a surrounding sea of particles that are bumping into it and forcing it to take on one classical state or another, rather than existing in a blurry in-between state. And the larger the physical system, the more quickly its quantum state decoheres into a classical state.

When particles of matter are pushed away from equilibrium by a flow of energy, or by a force like gravity, they form aggregates that decohere quickly, so the universe is constantly collapsing quantum reality and transforming it into classical reality. Decoherence explains why you don't need a conscious observer, or even a measuring device, to collapse wave functions: it happens naturally all the time so long as the quantum system is not isolated from its noisy environment.

But the concept of decoherence alone does not tell us why one particular measurement outcome occurs rather than another, and it does not answer the question of whether what happens is just a matter of chance. That's where *quantum Darwinism* comes in, a new and trendy interpretation of quantum mechanics that fits into the universal Darwinism paradigm perfectly, continuing the narrative that reality is fundamentally Darwinian. If the unifying theory of reality can explain both the fine-tuning problem and the measurement problem, it'll be more deserving of the title of theory of everything than any theory that has come before it.

## The Quantum-to-Classical Transition—A Form of Natural Selection

Quantum Darwinism says that the classical world emerges from the quantum world because most quantum systems are *open* systems that

interact with their surrounding environment, which causes decoherence. But the classical-scale outcome that emerges from those interactions is not selected randomly by nature. Specifically, a quantum state gets converted into a classical state that represents the most stable configuration between that quantum system and the systems it is interacting with in the environment. The collapse of the wave function is caused by a natural selection process known as *einselection*, which stands for *environment-induced superselection*.

Wojciech Zurek, the physicist at Los Alamos National Laboratory who pioneered quantum Darwinism, says that the classical state of the system that is selected from the menu of possibilities described by the wave function is the state that has the lowest internal entropy, which is also the state that produces the most environmental entropy. These preferred states are called *pointer states*, and they represent the information from the wave function that has survived the environmentally induced quantum collapse. In other words, the quantum-to-classical transition is a form of natural selection, and the classical world presumably evolves toward the most stable, organized world out of all possible worlds.

According to Zurek, the state of the system that survives the collapse is also the one that is best at imprinting its information on the environment. That is another way of saying that the selected system is the one that is most correlated with its environment and best adapted to its surroundings. As covered by science journalist Philip Ball for the digital magazine *Quanta*, recent experimental evidence supports quantum Darwinism's account of the wave function collapse, though more studies are needed to further test the theory.

From the "computational universe" perspective championed by Seth Lloyd, you could see the universe as a giant quantum computer that is using the computational power of quantum superposition to calculate the world trajectory that on average increases complexity and functionality. In this model, there is both *chance and necessity*: that is, reality has a probabilistic or stochastic nature, but outcomes are statistically biased toward goal states that can be described by attractors.

The gravity-induced formation of non-equilibrium structures such as stars and planets, which are both necessary requirements for universes with life as we know it, can be seen as emergences that occur as a result of the

Darwinian selection of certain possible quantum states over others. If this story is true, then Einstein was more right than wrong when he said, "God does not play dice." God *does* play dice, but *the game is rigged*. The trajectory of cosmic evolution that emerges from all of those probabilistic interactions is predictable—not in detail but in outline. All roads would seem to inevitably lead to the emergence of a cosmic mind, as the universe computes the teleological path through the realm of the possible.

Before moving on from quantum mechanics, I should note that in the model of reality proposed by poetic meta-naturalism, where there are true emergences in which new causal entities come into existence, there may be room for more than one interpretation of quantum mechanics. When information-processing agents emerge, the rules may work a little differently. In a multilevel model of reality, quantum Darwinism could describe the logic of the wave function collapse prior to the emergence of agents, while another interpretation could explain agent-induced wave function collapse. Two approaches to solving the quantum mechanical mystery that account for the actions of agents are *quantum Bayesianism* and *relational quantum mechanics*. Relational quantum mechanics, devised by Carl Rovelli, says that classical reality emerges from the correlations that get built up between interacting systems, like quantum systems and their environment.

As we learned in Chapter Six, Rovelli showed that Darwinian adaptation by natural selection builds up correlations between a system and its environment, which produces the adaptive information in a system that creates biological agency. Perhaps it is possible that Zurek's and Rovelli's solutions to the measurement problem may be converging on the same fundamental process. *Quantum Bayesianism*, or *QBism* for short, is another interpretation of quantum mechanics that is philosophically aligned with the new cosmic narrative presented in this book, and it is unique in that it talks about the role of experience. Its founder, Christopher Fuchs, describes QBism's solution to the measurement problem this way:

> In QBism, a measurement is an action an agent takes to elicit an experience. The measurement outcome is the experience so elicited. The measurement outcome is thus personal to the agent who takes the measurement action. In this sense, quantum mechanics, like

> probability theory, is a single user theory. A measurement does not reveal a pre-existing value. Rather, the measurement outcome is created in the measurement action.[8]

While it is not obvious how these different interpretations could all be part of an underlying explanation that connects the wave function collapse to cosmic evolution and knowledge creation, it would not be surprising to discover consilience among them. The unifying theory of reality, which says reality is a story of cosmic awakening, has the power to reveal such consilience. Together, cosmological natural selection and quantum Darwinism extend the integrated evolutionary synthesis so that the doctrine explains natural phenomena from the smallest to largest scales. Physicists couldn't see the new cosmic narrative because they were only looking at parts and not wholes, particles and not processes. The precise values of the fundamental constants are what they need to be for the Darwinian algorithm of variation and selection to work.

## What Does a Self-Organizing Universe Mean for Humanity?

Since the cosmic teleology paradigm suggests that the destiny of the universe and the destiny of intelligent life are intimately intertwined, it has some interesting implications for ethics and morality. Most contemporary philosophers would argue that any purpose we attribute to life is inherently subjective and therefore not real in any objective sense. The common assumption is that you "can't get an 'ought' from an 'is,'" meaning logically we cannot derive moral lessons from any facts about nature. But this is based on the myopic and parochial notion of a universe that is indifferent to life.

Because the universe has a built-in tendency to self-organize into an increasingly complex state, a collective purpose for intelligent life emerges from fundamental laws of nature. As humans, we are to a degree wired to see ourselves as something separate and distinct from nature, but this intuition is deceiving. We are, in our very essence, complex manifestations

of nature and a testament to the physical laws that drive the continual creation of novelty, both in form and function. If achieving a maximally complex state is nature's intrinsic objective, then it logically follows that sentient life, self-aware manifestations of nature that behave with intent, should strive to generally act in accordance with this innate tendency—something we've been genetically programmed to do long before anyone became cognizant of it, due to natural selection. It is the teleological nature of the universe that creates a cosmic context that instills life with meaning and purpose, and that purpose is to suffuse the cosmos with intelligence and experience by thriving, multiplying, and expanding.

As stated in the introduction, while the laws and constants of physics seem to have carved out this evolutionary trajectory for life in the universe, the hard truth is that existential success is in no way guaranteed for *Homo sapiens*. There is no particular law or force of nature preventing our specific civilization from destroying itself. The fact that humans are, at least to some degree, autonomous agents with agency and volition means that it is up to us to decide between extinction and transcendence, but we have every incentive to work toward the latter and away from the former. Through the collective efforts of humanity, by way of our intellectual, cultural, and technological progress, we can continue to assist the cosmos in its great awakening process. Promoting an awareness of our emergent cosmic purpose will facilitate exponential technological and social progress that can allow life to continually progress and the biosphere to extend itself, out into the cosmos.

## The Road to Omega

Such a cosmic agenda establishes clear milestones for humanity and a rational ethical code. This code is based on the premise that we have a responsibility to do what is good for the living network, which includes not only humans but all sentient life, as well as the biosphere, since life critically depends on entire ecosystems for survival. If we care about the futures of our children, and their children, such a task is a moral obligation. If we want to see a universe that isn't cold and lifeless, but increasingly sentient and creative, then it is also a spiritual obligation.

Many assume that when you get down to the nuts and bolts of nature, a spiritual worldview is simply incompatible with a scientific one. While that is the common assumption, it couldn't be more wrong. Spirituality simply refers to a sense of connection to something larger than oneself, and it has nothing to do with the supernatural. To quote Carl Sagan again, "Science is not only compatible with spirituality; it is a profound source of spirituality."[9] Einstein not only shared this view but also believed that the only path to true spiritual enlightenment was the trail we travel along when we pursue scientific discovery. "The further the spiritual evolution of mankind advances, the more certain it seems to me that the path to genuine religiosity does not lie through the fear of life, and the fear of death, and blind faith, but through striving after rational knowledge."[10]

So, what is the meaning of life, and the meaning of existence, according to this new worldview? In his book *Vital Dust*, Nobel laureate Christian de Duve gives a satisfying answer to that question:

> If the universe is not meaningless, what is its meaning? For me, this meaning is to be found in the structure of the universe, which happens to be such as to produce thought by way of life and mind. Thought, in turn, is a faculty whereby the universe can reflect upon itself, discover its own structure, and apprehend such immanent entities as truth, beauty, goodness, and love. Such is the meaning of the universe, as I see it.

But to realize its meaning and transcendent purpose, the cosmos needs our help. As James Gardner puts it in his book *Biocosm*: "The universe, de Duve would surely agree, cannot exhibit the miracle of conscious thought nor its finest manifestations—truth, beauty, goodness and love—by itself. To do so, the universe urgently needs the assistance of humble mortals like ourselves."

So, what can we do right now to assist the universe in its process of cosmic awakening?

Beyond self-awareness, there is a higher level of awareness that can be attained by intelligent agents called *meta-awareness*. This new level of self-reference would refer to awareness of self-awareness, but it would also

indicate that the agent understands that it is part of a larger network of agents that are all part of an interdependent, interconnected whole that we call the biosphere, and more specifically, the global brain. Behaviorally, meta-awareness means you are aware of your causal power and choose to optimize your decision making so that your choices are consistent with your long-term goals and the goals of the society one is embedded in.

Meta-awareness must be consciously cultivated in order for it to fully emerge in an agent, and when awareness of awareness of awareness goes mainstream, the resulting synergy will take humanity to new heights. But meta-awareness is not a luxury; if we do not become cognizant of the fact that the only way we can overcome our collective existential challenges is by working together, our civilization will fail. Therefore, while nations must retain their individual identities—remember, there is strength in diversity—they must also align common interests, which produces synergy through minimizing conflict and promoting cooperation. Optimal complexity and computational capacity comes from a balance of diversity or differentiation and integration or connection.

To align interests, we must have a common worldview, and since "cosmic teleology" is a loaded term, perhaps a better name for the worldview presented in this book is the "cosmic perspective," which is not unlike the religion envisioned by Sagan and Einstein. Sagan said, "A religion, old or new, that stressed the magnificence of the Universe as revealed by modern science might be able to draw forth reserves of reverence and awe hardly tapped by the conventional faiths. Sooner or later such a religion will emerge."[11]

"Religion" is another a loaded term, but a spiritual ideology that is guided by science and aided by technology, that has a universal morality and a shared existential goal, will be the worldview of the future. It has to be if we want our civilization to survive. Under the cosmic perspective, there is no "us versus them," there's only "we." Since we're all part of an interdependent whole, our goal should be to try to achieve *the greatest good for the greatest number of people*.

While the dream of a cosmic religion might sound hopelessly idealistic to some, I do believe it is a supremely reasonable and attainable worldview, one that any enlightened society will eventually move in the direction of. A

2017 study that surveyed over six hundred people found that psychedelic experiences, by dissolving one's ego and ideological framework, reliably shifted people's political beliefs and attitudes toward ones that could be described as more compassionate and progressive.[12]

Specifically, individuals became more opposed to authoritarianism and more concerned with the well-being of others and of nature. So, things like scientific enlightenment and higher states of consciousness naturally bring us closer to the cosmic perspective by revealing the interconnected nature of our society. I'm not suggesting that everyone over the age of eighteen eat magic mushrooms tonight, although that would be a fascinating experiment. I am saying we need to have a cognitive reset, and we can do that simply by thinking of ourselves as global citizens of the planetary superorganism called Gaia.

Our primary collective goal should be hyperconnection. Every year the number of ways to socially connect with others, even those in the most remote locations, increases exponentially, allowing for ever-greater information flow. This inevitably promotes greater social awareness and knowledge exchange. It is also a mathematical fact that increasing the number of connections between nodes increases the computational power of a network. If human civilization really is one massive neural network now, with each individual and device acting as a node or neuron, then that means social media has the ability to create some powerful network effects. We must utilize all the tools of the information age—and invent new ones. The power of blockchain technology to facilitate the emergence of self-organizing systems of all types is just beginning to be explored.

Yes, the world has some serious problems, but if we did not have problems, we would never be forced to find new solutions. Problems push progress forward. Let's embrace our ultimate existential challenges and come together to solve them. It is time to forget our differences and think of ourselves only as humans, engaged in a common biological and moral struggle. If the cosmic perspective, and the philosophy of poetic meta-naturalism, or some similar worldview of evolution and emergence, can build a bridge between the reductionist worldview and the religions of the world, then we can be optimistic that a new level of order and functionality will emerge from the current sea of chaos.

Knowledge is enlightenment, knowledge is transcendence, and knowledge is power. The tendency toward disorder described by the second law requires that life acquire knowledge forever, giving us all an individual and collective purpose by creating the constraint that forces us to create. By becoming aware of our emergent purpose, we can live more meaningful lives, in harmony with one another and with the aspirations of nature. You are not a cosmic accident. You are a cosmic imperative.

# NOTES

## Chapter 1


1. Thomas Henry Huxley, *Evidence as to Man's Place in Nature* (Williams & Norgate: 1863).
2. Francis Crick, *Life Itself: Its Origin and Nature* (Simon & Schuster: 1981).
3. Carl Sagan, "The Abundance of Life-Bearing Planets," *Bioastronomy News* 7, no. 4 (1995): 1–4
4. Christian de Duve, *A Guided Tour of The Living Cell* (W. H. Freeman & Co.: 1984).


## Chapter 3


1. Ilya Prigogine, *The End of Certainty: Time, Chaos and the New Laws of Nature* (Free Press: 1997).
2. Elan Stopnitzky and Susanne Still, "Non-equilibrium Odds for the Emergence of Life," *arXiv*, arXiv:1705.02105 [physics.bio-ph], https://arxiv.org/abs/1705.02105v1


## Chapter 4


1. Harold Morowitz and Eric Smith, "Energy flow and the organization of life," *Complexity* 13, no. 1 (October 2007): 51–59, doi:10.1002/cplx.20191.
2. Sara Walker, "The Origin and Nature of Life on Earth: The Emergence of the Fourth Geosphere," *Physics Today* 60, no. 9 (September 2017): 58, doi:10.1063/PT.3.3694.
3. Olivia P. Judson, "The Energy Expansions of Evolution," *Nature Ecology & Evolution* 1 (April 2017), doi:10.1038/s41559-017-0138.
4. M. J. Russell et al., *Natural and Laboratory-Simulated Thermal Geochemical Processes*, ed. R. Ikan, (2003), 325–388; W. Martin & M. J. Russell, Phil. Trans. Roy. Soc. B, online publication, doi:10.1098/rstb.2006.1881 (2006)
5. William James, *The Principles of Psychology* (Henry Holt and Company: 1890).

# Chapter 5


1. Edward C. McIrvine and Myron Tribus, "Energy and Information," *Scientific American* 225, no. 3 (September 1971): 179–190.
2. "John Maynard Smith—20th Century Biology as a 'Science of Information' (95/102)," https://www.youtube.com/watch?v=78ikxE5-POY.
3. Artemy Kolchinsky and David H. Wolpert, "Semantic Information, Autonomous Agency, and Non-Equilibrium Statistical Physics," *Interface Focus* 8, no. 6 (December 2018), doi:10.1098/rsfs.2018.0041.


# Chapter 6


1. David Deutsch, "David Deutsch on the infinite reach of knowledge," TED interview, https://www.ted.com/talks/the_ted_interview_david_deutsch_on_the_infinite_reach_of_knowledge/up-next?language=en.
2. Daniel Dennett (@danieldennett), Twitter, October 15, 2019, 4:59 p.m., https://twitter.com/danieldennett/status/1184212103969869825.
3. Freeman J. Dyson, "Energy in the Universe," *Scientific American* (September 1971).
4. Pietro Gori, *Nietzsche's Pragmatism: A Study on Perspectival Thought* (De Gruyter: 2020), 10.
5. Karl Popper, "Evolutionary Epistemology," in *Open Questions in Quantum Physics*, ed. Gino Tarozzi and Alwyn van der Merwe (Springer: 1985): 395–413.
6. Ibid.
7. Ibid.
8. Konrad Lorenz, *Behind the Mirror: A Search for a Natural History of Human Knowledge* (Harcourt, Brace: 1977).
9. Terrence W. Deacon, "Shannon—Boltzmann—Darwin: Redefining Information (Part I)," *Cognitive Semiotics* (September 2007), doi:10.1515/cogsem.2007.1.fall2007.123.
10. Richard Dawkins, *The Selfish Gene* (Oxford University Press: 1976).
11. Karl Friston, "The mathematics of mind-time," Aeon, May 2017, https://aeon.co/essays/consciousness-is-not-a-thing-but-a-process-of-inference.
12. François Jacob, *Of Flies, Mice, and Men* (Harvard University Press: 1998).
13. David Deamer, *First Life: Discovering the Connections Between Stars, Cells, and How Life Began* (University of California Press: 2011).
14. Terrence W. Deacon, "Shannon—Boltzmann—Darwin: Redefining information (Part II)," *Cognitive Semiotics* (2007), doi:10.3726/81600_123.
15. Carlo Rovelli, "Meaning and Intentionality = Information + Evolution," in *Wandering Towards a Goal: How Can Mindless Mathematical Laws Give Rise to Aims and Intention?* (Springer: 2018).
16. Ibid.
17. Toby Shannan and David Krakauer, "David Krakauer," 2021, in *Ideas at Work with Toby Shannan*, https://podcasts.apple.com/us/podcast/ep-5-david-krakauer/id1509519346?i=1000477690144.

# Chapter 7


1. "Brian Greene: The Search for Meaning in an Evolving Universe," *New Scientist*, https://www.youtube.com/watch?v=-o0TJy67Gi8.
2. Bloggingheads.tv, "Science Saturday: A Plan to Dye One's Whiskers Green | John Horgan & David Deutsch," April 5, 2018, https://www.youtube.com/watch?v=y3Afrwe-Yyg&t=2s.
3. Sara Walker, "The Descent of Math," in *Trick of Truth: The Mysterious Connection Between Physics and Mathematics?*, ed. A. Aguirre, B. Foster, and Z. Merali (Springer: 2016).
4. TED, "Alex Wissner-Gross: A New Equation for Intelligence," February 6, 2014, https://www.youtube.com/watch?v=ue2ZEmTJ_Xo&t=329s.
5. "Brian Greene: The Search for Meaning in an Evolving Universe," *New Scientist*, April 7, 2020, https://www.youtube.com/watch?v=-o0TJy67Gi8.
6. Freeman J. Dyson, "Time Without End: Physics and Biology in an Open Universe," *Reviews of Modern Physics* 51, no. 3 (July 1979).
7. BrightSight Speakers, "Sean Carroll: God Is Not a Good Theory," June 10, 2015, https://www.youtube.com/watch?v=Vs8seDbzNl8.
8. David Deutsch, *The Beginning of Infinity: Explanations That Transform the World* (Penguin Books: 2012).
9. Freeman J. Dyson, "Time Without End: Physics and Biology in an Open Universe," *Reviews of Modern Physics* 51, no. 3 (July 1979).
10. Seth Lloyd, "The Computational Universe," in *Information and the Nature of Reality: From Physics to Metaphysics*, ed. P. Davies and N. H. Gregersen (Cambridge University Press: 2014), 125.
11. Ray Kurzweil, *The Singularity Is Near* (Viking: 2005).
12. Ibid.
13. Richard Dawkins, "Review: Human Chauvinism," *Evolution* 51, no. 3 (June 1997), doi:10.2307/2411179.
14. S. J. Gould, "On replacing the idea of progress with an operational notion of directionality," *Evolutionary Progress* (University of Chicago Press: 1988), 319–338.


# Chapter 8


1. E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press: 2003).
2. Karl Friston, "The Mathematics of Mind-Time," Aeon, May 2017, https://aeon.co/essays/consciousness-is-not-a-thing-but-a-process-of-inference.
3. Steven Pinker, "The Second Law of Thermodynamics," Edge, 2017, https://www.edge.org/response-detail/27023.
4. Adrian Bejan, *Design in Nature: How the Constructal Law Governs Evolution in Biology, Physics, Technology, and Social Organization* (Doubleday: 2012).
5. *Frontiers*, "Empowering Robots for Ethical Behavior," *ScienceDaily* (July 2017).
6. John Naughton, "Ashby's Law of Requisite Variety," Edge, 2017, https://www.edge.org/response-detail/27150.
7. Robin L. Carhart-Harris et al., "The Entropic Brain: A Theory of Conscious States Informed by Neuroimaging Research with Psychedelic Drugs," *Frontiers in Human Neuroscience* (February

2014), doi:10.3389/fnhum.2014.00020.
8. Christof Koch, "Is Consciousness Universal?," *Scientific American* (January 2014).
9. F. Heylighen and C. Joslyn, "The Law of Requisite Variety," Principia Cybernetica Web, August 31, 2001 (modified August 31, 2001), http://pespmc1.vub.ac.be/REQVAR.html.
10. Eric J. Chaisson, *Cosmic Evolution: The Rise of Complexity in Nature* (Harvard University Press: 2002).
11. Roger Lewin, *Complexity: Life at the Edge of Chaos* (Collier: 1993).

# Chapter 9

1. Paul Davies, *The Cosmic Blueprint* (Simon & Schuster: 1989).
2. "Self-organization in cybernetics," Wikipedia, accessed December 21, 2021, https://en.wikipedia.org/wiki/Self-organization_in_cybernetics.
3. Harold J. Morowitz, *The Emergence of Everything: How the World Became Complex* (Oxford University Press: 2004).
4. Paul Davies, *The Fifth Miracle: The Search for the Origin and Meaning of Life* (Simon & Schuster: 2000).
5. Ernst Mayr, "The Objects of Selection," *PNAS* 94, no. 6 (March 1997): 2091–2094, doi:10.1073/pnas.94.6.2091.
6. Erik P. Hoel, "Agent Above, Atom Below: How Agents Causally Emerge from Their Underlying Microphysics," in *Wandering Towards a Goal: How Can Mindless Mathematical Laws Give Rise to Aims and Intention?* (Springer: 2018).
7. Susanne Still et. al, "The Thermodynamics of Prediction," *arXiv*, doi:10.1103/Phys RevLett.109.120604.
8. Roger Lewin, *Complexity: Life at the Edge of Chaos* (Collier: 1993).

# Chapter 10

1. Bloggingheads.tv, "Science Saturday: A Plan to Dye One's Whiskers Green | John Horgan & David Deutsch," April 5, 2018, https://www.youtube.com/watch?v=y3Afrwe-Yyg&t=2s.
2. Jerry A. Fodor, *A Theory of Content and Other Essays* (Bradford Books: 1992).
3. Christof Koch, "Is Consciousness Universal?", *Scientific American* (January 2014).

# Chapter 11

1. "Reflections on Free Will: A Review by Daniel C. Dennett," samharris.org, January 26, 2014, https://www.samharris.org/blog/reflections-on-free-will.
2. Sara Walker, "Bio from Bit," in *Wandering Towards a Goal: How Can Mindless Mathematical Laws Give Rise to Aims and Intention?* (Springer: 2018).
3. U. Maoz et al., "Neural Precursors of Decisions That Matter—An ERP Study of Deliberate and Arbitrary Choice," *eLife* (October 2019), doi:107554/eLife.39787.

4. Steve Paulson, "The Spiritual, Reductionist Consciousness of Christof Koch," *Nautilus*, April 6, 2017, https://nautil.us/issue/47/consciousness/the-spiritual-reductionist-consciousness-of-christof-koch.

# Chapter 12

1. Ned Block, "On a confusion about a function of consciousness," *Behavioral and Brain Sciences* 18, no. 2 (June 1995), 227–247, doi:10.1017/S0140525X00038188.
2. Adam Safron, interview by the author over Zoom, July 10, 2021.
3. U. Maoz et al., "Neural Precursors of Decisions That Matter—An ERP Study of Deliberate and Arbitrary Choice," *eLife* (October 2019), doi:107554/eLife.39787.
4. Simon DeDeo, "Mutual Explainability, or, a Comedy in Computerland," FQXi, April 19, 2020, https://fqxi.org/community/forum/topic/3492.

# Chapter 13

1. Christof Koch, "Is Consciousness Universal?," *Scientific American* (January 2014).
2. Paul Davies, *The Goldilocks Enigma: Why Is the Universe Just Right for Life?* (Houghton Mifflin: 2008).
3. Stephen Hawking, *A Brief History of Time* (Bantam Books: 1998).
4. raschau, "Richard Dawkins on Deism: Has Science Buried God? (2011)," January, 19, 2019, https://www.youtube.com/watch?v=jBOGfFTMlNo.
5. Freeman J. Dyson, *Infinite in All Directions* (HarperCollins: 2004).
6. Daniel C. Dennett, *Darwin's Dangerous Idea: Evolution and the Meanings of Life* (Simon & Schuster: 1996).
7. Lee Smolin, "Black holes in cosmological natural selection," *Physics Today* 67, no. 10 (2014), doi:10.1063/PT.3.2529.
8. Christopher Fuchs and Ruediger Schack, "QBism and the Greeks: Why a quantum state does not represent an element of physical reality," *Physics Scripta* 90, no. 1 (December 2014), doi:10.1088/0031-8949/90/1/015104.
9. Carl Sagan, *The Demon-Haunted World: Science as a Candle in the Dark* (Ballantine Books: 1997).
10. Albert Einstein, *Out of My Later Years* (Gramercy: 1993).
11. Carlo Rovelli, "Meaning and Intentionality = Information + Evolution," in *Wandering Towards a Goal: How Can Mindless Mathematical Laws Give Rise to Aims and Intention?* (Springer: 2018).
12. Matthew M. Nour, Lisa Evans, and Robin L. Carhart-Harris, "Psychedelics, Personality and Political Perspectives," *Journal of Psychoactive Drugs* 49, no. 3 (April 2017): 182–191, doi:10.1080/02791072.2017.1312643.

# ABOUT THE AUTHOR

Bobby Azarian is a science journalist and a cognitive neuroscientist with a PhD from the Krasnow Institute for Advanced Study at George Mason University. He has written for publications including *The Atlantic, The New York Times*, BBC Future, *Scientific American*, Slate, HuffPost, and Aeon, and his research has been published in peer-reviewed journals such as *Human Brain Mapping, Cognition & Emotion*, and *Acta Psychologica*. His blog *Mind in the Machine*, hosted by *Psychology Today*, has received more than eight million views. Azarian worked on season two of the YouTube Premium psychology-based series *Mind Field* (as a consultant and researcher), which helped the show win its first Emmy nomination. He is based in Arlington, Virginia.